

**Stochastic Mapping for Chemical Plume Source
Localization with Application to Autonomous
Hydrothermal Vent Discovery**

by

Michael V. Jakuba

B.S., Massachusetts Institute of Technology (2000)

M.S., Massachusetts Institute of Technology (2003)

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

and

WOODS HOLE OCEANOGRAPHIC INSTITUTION

February 2007

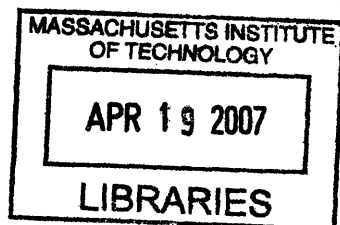
© Michael V. Jakuba, 2007. All rights reserved. The author hereby
grants to MIT and WHOI permission to reproduce and distribute
publicly paper and electronic copies of this thesis in whole or in part.

Author
Joint Program in Oceanography/Applied Ocean Science &
Engineering
January 7, 2007

Certified by
Dana R. Yoerger
Associate Scientist, WHOI
Thesis Supervisor

Accepted by
Lallit Anand
Chair, Graduate Committee

Accepted by
Henrik Schmidt
Chair, Joint Committee for Applied Ocean Science & Engineering



ARCHIVES

Stochastic Mapping for Chemical Plume Source Localization with Application to Autonomous Hydrothermal Vent Discovery

by

Michael V. Jakuba

Submitted to the Joint Program in Oceanography/Applied Ocean Science &
Engineering on January 9, 2007,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Abstract

This thesis presents a stochastic mapping framework for autonomous robotic chemical plume source localization in environments with multiple sources. Potential applications for robotic chemical plume source localization include pollution and environmental monitoring, chemical plant safety, search and rescue, anti-terrorism, narcotics control, explosive ordinance removal, and hydrothermal vent prospecting. Turbulent flows make the spatial relationship between the detectable manifestation of a chemical plume source, the plume itself, and the location of its source inherently uncertain. Search domains with multiple sources compound this uncertainty because the number of sources as well as their locations is unknown a priori.

Our framework for stochastic mapping is an adaptation of occupancy grid mapping where the binary state of map nodes is redefined to denote either the presence (occupancy) or absence of an active plume source. A key characteristic of the chemical plume source localization problem is that only a few sources are expected in the search domain. The occupancy grid framework allows for both plume detections and non-detections to inform the estimated state of grid nodes in the map, thereby explicitly representing explored but empty portions of the domain as well as probable source locations. However, sparsity in the expected number of occupied grid nodes strongly violates a critical conditional independence assumption required by the standard Bayesian recursive map update rule. While that assumption makes for a computationally attractive algorithm, in our application it results in occupancy grid maps that are grossly inconsistent with the assumption of a small number of occupied cells. To overcome this limitation, several alternative occupancy grid update algorithms are presented, including an exact solution that is computationally tractable for small numbers of detections and an approximate recursive algorithm with improved performance relative to the standard algorithm but equivalent computational cost.

Application to hydrothermal plume data collected by the autonomous underwater vehicle ABE during vent prospecting operations in both the Pacific and Atlantic oceans verifies the utility of the approach. The resulting maps enable nested surveys for homing-in on seafloor vent sites to be carried out autonomously. This eliminates inter-dive processing, recharging of batteries, and time spent deploying and recovering the vehicle that would otherwise be necessary with survey design directed by human operators.

Thesis Supervisor: Dana R. Yoerger
Title: Associate Scientist, WHOI

Acknowledgments

Graduate school has been a long but ultimately gratifying experience. There are many people without whom I simply would not have persevered.

Dana Yoerger, my adviser, I thank for the many opportunities provided, in particular for bringing me along on ABE cruises, and for the constant reminders to focus on real problems and the big picture. I still have much to learn and look forward to a continuing collaboration. My committee members Hanu Singh and Warren Seering have been tremendously helpful and accomodating, particularly in travelling between Woods Hole and Cambridge to attend committee meetings, I wish to thank Hanu in addition for the many thoughtful discussions on my work and otherwise, for providing access to the SeaBED AUV, and finally for handling an unexpected call from the Falmouth Waste Water Management Department with aplomb. I am also grateful to Chris German, Tim Shank, and Charlie Langmuir for providing opportunities to extend ABE's capabilities at sea. Rich Camilli kept my committee focused during the defense, and I thank him for his especially thorough reading of this thesis.

Working with the ABE group, especially at sea, has been the highlight of my studies in the Joint Program. Dana Yoerger makes it happen. Al Bradley is a consummate electrical engineer with understanding and insight that are often overwhelming. Rod Catanach is the only one who absolutely cannot screw up without catastrophic loss of the vehicle. ABE's two hundred dives and counting speak to his calm, methodical, careful, and supremely competent handling of ABE operations on ship and shore. Al Duester and Andy Billings are both irreplaceable members of the Autonomous Benthic Explorer (ABE) team.

My lab mates Chris, Oscar, Ryan, Matt, Cara, Ballard, and Gabriele have been collaborators, conspirators, commiserators, and motivators but foremost friends. Chris, Oscar, and Ryan's theses each set a very high bar. I also thank Ann Stone for ensuring I always had a paycheck, but mostly for always having a smile, even on the many Sunday afternoons in the office.

Julia and Marsha in the Academic Programs Office truly make the Joint Program home to its students. They are always supportive and accommodating, and they are largely responsible for the exceptional quality of life enjoyed by graduate students in the Joint Program.

Chris, Dirk, Oscar, Jim, Travis, Christian, and Rich have made for fine house mates and many adventures. The *Sea Pig* and friendships with those that helped build her and who rode her feeble wake will remain among my fondest memories of life on Millfield Street. Nick and Amy tell the best jokes. Steph saves birthday cakes and birthdays too. Leo, Tim, Will, Jonas, Dongo and Joe are like family and have helped keep the last six and a half years in perspective. I cannot imagine a better group of friends to go through life with.

My parents, Stan and Eva, have been unwavering in their support throughout my education. I value their pride in my accomplishments more than any degree. My sister Caroline is so much more impressive as an individual than she knows. I look forward to helping her through the inevitable rough spots in doctoral program she is presently applying to.

My fiancée Rachel has patience beyond measure, except on an empty stomach. Throughout she has believed in me more than I ever have in myself.

Thank you.

Mike

This work was supported in part by a National Science Foundation graduate research fellowship, a Presidential fellowship from the MIT Mechanical Engineering Department, WHOI Academic Programs, the National Science Foundation (grants OCE-0644290 and OCE-0241913), and the National Ocean and Atmospheric Administration (grant NA06OAR4600098).

Contents

1	Introduction	23
1.1	Background	24
1.1.1	Examples from the Natural World: Moths, Starfish, Lobsters and Crabs	25
1.1.2	Robotic Plume Source Localization	27
1.1.3	Hydrothermal Vent Prospecting	29
1.2	Approach	31
1.2.1	Nested Surveys for Chemical Plume Source Localization . . .	31
1.2.2	Occupancy Grid Mapping and its Application to Chemical Plume Source Localization	35
1.3	Contributions	36
1.4	Document Structure	38
I	Bayesian Occupancy Grid Mapping with Binary Mea- surements in Environments with Few Occupied Cells	41
2	Exact Occupancy Grid Mapping for Binary Measurements	43
2.1	Background: Occupancy Grids	44
2.1.1	Bayesian occupancy grid (OG) Mapping	45
2.1.2	Map Artifacts	49
2.1.3	Low Prior Environments	50
2.2	Forward and Inverse Sensor Models	53
2.3	An Invertible Forward Model for Binary Measurements	54
2.3.1	Inversion and Marginalization	56

2.3.2	Posterior Independence Property	59
2.4	An Exact Algorithm for the Marginal Posteriors	60
2.4.1	Numerical and Computational Considerations	63
2.5	Summary, Contributions, and Future Work	67
3	Approximate Algorithms for Low Prior Environments	69
3.1	Approximate Algorithms from Relaxed conditional independence of measurements (CIM) Assumptions	70
3.1.1	Standard Algorithm for Binary Measurements	74
3.1.2	Extended-Standard Algorithm	77
3.1.3	Conditional Independence of Detections Algorithm	79
3.2	Approximate Algorithms from Assuming Independence of the Posteriors	81
3.2.1	Independence of Posteriors (IP) Algorithm	82
3.3	Simulation	88
3.4	Summary, Contributions, and Future Work	101
II	Automated Nested Survey for Hydrothermal Vent Lo- calization and Mapping	104
4	Hydrothermal Plume Survey by Autonomous Underwater Vehicle	105
4.1	Hydrothermal Plumes from an autonomous underwater vehicle (AUV)'s Perspective	107
4.1.1	Buoyant Hydrothermal Plumes	108
4.1.2	Non-buoyant hydrothermal plumes	111
4.2	Background: A Three-Stage Nested Survey Approach	114
4.3	Anomaly Maps	117
4.3.1	A Taxonomy of Hydrothermal Tracers	119
4.3.2	Anomaly Definition	121
4.3.3	Preprocessing Applied	122
4.3.4	Background Profile	127
4.3.5	Declaring Detections	131
4.3.6	Results from Additional Dives	139
4.4	Water Velocity Profiles from On-Board ADCP	141

4.4.1	Error Sources	141
4.4.2	Effect of Heading Error	142
4.4.3	Results	148
4.5	Uncertainty in Seafloor Vent Location	153
4.5.1	Buoyant Plume Interception	154
4.5.2	Distribution of buoyant plume (BP) Detection Location	156
4.5.3	Dependence on source-relative vehicle height h_{vs}	157
4.6	Conclusions	170
5	Buoyant Hydrothermal Plume Source Localization via Nested Survey	175
5.1	Introduction	175
5.2	Implementation	176
5.2.1	Forward Model for BP Detection	177
5.2.2	Input Data	184
5.2.3	Mapping Domain	185
5.2.4	Choice of OG Algorithm	186
5.3	Results: Vent Field OG Maps	187
5.3.1	Conversion from Field Map to Vent Map	188
5.3.2	Survey Design from OG Map	191
5.4	Conclusions	194
6	Conclusion	195
A	Notation	203
B	Exact Binary Occupancy Grid Mapping Derivations	207
B.1	Sums Over Maps	207
B.2	Derivation: Binary Inverse Model	208
B.3	Proof: $p(\mathbf{m} \mid \bar{D}^t) = \prod_{s \in \mathcal{S}} \bar{P}_s^t \cdot \prod_{s \in \bar{\mathcal{S}}} (1 - \bar{P}_s^t)$	209
B.4	Proof: $P[\bar{d}^t \mid \bar{D}^{t-1}, m_c] \neq P[\bar{d}^t \mid m_c]$	212
B.5	Derivation: Exact Solution	214
C	Exact OG mapping algorithm with unconnected observation regions	219

D	IPAlgorithm Derivations	225
D.1	Recursive IP	225
D.2	Extended IP	227
E	Connections to Pang, 2004	229
F	Anomaly Maps	237
F.1	ELSC: Site-1 (Kilo Moana)	239
F.1.1	ABE-126	241
F.1.2	ABE-137	248
F.2	ELSC: Site-3 (ABE Site)	253
F.2.1	ABE-128	254
F.2.2	ABE-136	261
F.3	ELSC: Site-5	266
F.3.1	ABE-131	267
F.3.2	ABE-133	274
F.3.3	ABE-134	279
F.4	SMAR	283
F.4.1	ABE-150	285
F.4.2	ABE-151	292
F.4.3	ABE-153	297
G	Comparison of OG Algorithms for Various Prior	303
G.1	Results: ABE-151 at Nominal Prior	304
G.2	Algorithm Performance for Various Priors	304

List of Figures

1-1	The nested survey approach to chemical plume source localization. . .	32
1-2	A spectrum of chemical plume source localization methodologies differentiated by the strength of coupling between vehicle trajectory and sensory input.	33
2-1	Schematic depiction of regions likely to contain an upwind source. . .	51
2-2	Candidate explanations for two plume detections.	51
2-3	Schematic sonar range-finder decomposed into a binary detection and non-detection.	56
2-4	Inverse sensor model entropy	58
2-5	Standard and exact OG mapping compared on a two-cell map. . . .	64
2-6	Computation time for our exact OG mapping algorithm versus number of detections and map size.	65
2-7	Numerical failure of the exact algorithm in a two-cell map.	66
3-1	Recursive update of a 4-cell map via the standard and recursive IP algorithms versus the exact result.	84
3-2	Comparison of OG maps from a deterministic sequence of $n^t = 5$ successive detections.	91
3-3	The OG maps of Fig. 3-2, but after $n^t = 10$ successive detections. . .	92
3-4	Occupancy grid maps generated in simulation by four approximate OG mapping algorithms after 150 measurements.	93
3-5	The maps of Fig. 3-4 after 300 measurements.	94
3-6	The maps of Figs. 3-4 & 3-5 after 1000 measurements.	95
3-7	Error between the actual and expected number of occupied cells for maps produced using the IP algorithm.	97

3-8	Average map score vs. time for various revisited map fractions. . . .	100
3-9	Expected numbers of occupied cells within selected regions of an OG map.	101
3-10	OG algorithm choice based on map prior and expected number of detections.	103
4-1	In situ temperature records collected by ABE as it passed through buoyant plumes at several different heights above the seafloor. . . .	109
4-2	Optical backscatter data collected by ABE in a non-buoyant plume adjusted for advection.	112
4-3	The mid-ocean ridge system, and the locations of two recent hydrothermal vent-finding expeditions that utilized an AUV as complementary to towed assets. ELSC (20 °3' S, 176 °12' W): Eastern Lau Spreading Center (September/October 2004). SMAR (4 °54' S, 12 °28' W): Southern Mid-Atlantic Ridge (March 2005).	114
4-4	Vehicle tracklines from three nested surveys at a hydrothermal site. .	116
4-5	Potential temperature and OBS profiles showing contamination by a NBP.	130
4-6	Background profiles of potential temperature and optical backscatter from vehicle descent.	130
4-7	Optical backscatter and potential temperature anomaly following removal of background profile.	131
4-8	Tracer time-series following detrending for buoyant plume detection. .	132
4-9	Potential temperature anomaly observed during ABE-126.	135
4-10	Optical backscatter anomaly observed during ABE-126.	136
4-11	Anomalies classified by a Hampel Identifier as detections during ABE-126.	138
4-12	Effect of heading error on earth-frame water velocity estimates. . . .	144
4-13	Schematic dependence of water velocity error on crabbing angle. . . .	146
4-14	Non-dimensional error in water velocity magnitude and direction from heading error as a function of crabbing angle.	147
4-15	Crossflow speed and direction profiles vs. time from ABE-151.	149
4-16	Birds-eye view of crossflow speed and direction observed during ABE-151.	150

4-17	LBL-derived crossflow velocity.	151
4-18	LBL-derived crossflow velocity: loss of temporal resolution relative to DVL-derived crossflow	152
4-19	Buoyant plume interception (perspective view)	155
4-20	The effect of ambient crossflow strength U_0 on the location of a plume at vehicle depth assuming constant rise rate W_0 : (a) strong crossflow ($U_0/W_0 > 1$); (b) weak crossflow ($U_0/W_0 < 1$). Higher intensity crossflows advect the plume over larger horizontal distances for the same vertical source-vehicle separation h_{vs} . Within the plume, the average distance to the plume centerline $E[\mathbf{r}]$ is less than the average distance to the plume source on the seafloor $E[\mathbf{r}_0]$; however, wide plumes and weak advection reduce the magnitude of the difference.	156
4-21	Distance to plume centerline vs. distance to source, simulated data.	158
4-22	Measured distance to source versus vehicle height.	160
4-23	Buoyant plume detections: distance to plume centerline versus distance to source.	168
4-24	Attainable reduction in source location ambiguity from on-board mea- surement of crossflow.	171
5-1	OG mapping applied to nested survey for hydrothermal vent localization.	178
5-2	Observed plume width vs. height.	182
5-3	OG Map produced from Phase-1 dive ABE-128.	189
5-4	OG Map produced from Phase-2 dive ABE-136 using the map of Fig. 5- 3 to define the prior.	190
5-5	OG map of Fig. 5-4 transformed from probability of vent field occu- pancy into individual vent occupancy.	192
5-6	Survey coverage relative to high and low value map area identified in OG maps.	193
C-1	Schematic definitions of dependent portions of an OG map after mul- tiple detections.	220
C-2	Schematic depicting the independent update of map cells in disjoint subsets of detection regions.	223

E-1	Pang's [101,102] algorithm vs. the recursive IP and exact OG mapping algorithms in a single-source domain.	234
E-2	Pang's [101,102] algorithm vs. the recursive IP and exact OG mapping algorithms in a multiple-source domain.	235
F-1	Vent prospecting summary data, ELSC-1.	239
F-2	Closeup of vent prospecting summary data, ELSC-1.	240
F-3	Descent profiles vs. depth, ABE-126.	241
F-4	Descent profiles vs. potential density, ABE-126.	241
F-5	Time series of hydrothermal tracers after pre-processing, ABE-126.	242
F-6	Time series of hydrothermal tracers processed for NBP detection, ABE-126.	243
F-7	Time series of hydrothermal tracers processed for BP detection, ABE-126.	244
F-8	Bird's eye view of anomaly intensity after processing for NBP detection, ABE-126.	245
F-9	Bird's eye view of anomaly intensity after processing for BP detection, ABE-126.	246
F-10	Bird's eye view of ADCP-derived crossflow velocity, ABE-126.	247
F-11	Descent profiles vs. depth, ABE-137.	248
F-12	Descent profiles vs. potential density, ABE-137.	248
F-13	Time series of hydrothermal tracers after pre-processing, ABE-137.	249
F-14	Time series of hydrothermal tracers processed for BP detection, ABE-137.	250
F-15	Bird's eye view of anomaly intensity after processing for BP detection, ABE-137.	251
F-16	Bird's eye view of ADCP-derived crossflow velocity, ABE-137.	252
F-17	Vent prospecting summary data, ELSC-3.	253
F-18	Descent profiles vs. depth, ABE-128.	254
F-19	Descent profiles vs. potential density, ABE-128.	254
F-20	Time series of hydrothermal tracers after pre-processing, ABE-128.	255
F-21	Time series of hydrothermal tracers processed for NBP detection, ABE-128.	256

F-22 Time series of hydrothermal tracers processed for BP detection, ABE-128.	257
F-23 Bird's eye view of anomaly intensity after processing for NBP detection, ABE-128.	258
F-24 Bird's eye view of anomaly intensity after processing for BP detection, ABE-128.	259
F-25 Bird's eye view of ADCP-derived crossflow velocity, ABE-128.	260
F-26 Descent profiles vs. depth, ABE-136.	261
F-27 Descent profiles vs. potential density, ABE-136.	261
F-28 Time series of hydrothermal tracers after pre-processing, ABE-136.	262
F-29 Time series of hydrothermal tracers processed for BP detection, ABE-136.	263
F-30 Bird's eye view of anomaly intensity after processing for BP detection, ABE-136.	264
F-31 Bird's eye view of ADCP-derived crossflow velocity, ABE-136.	265
F-32 Vent prospecting summary data, ELSC-5.	266
F-33 Descent profiles vs. depth, ABE-131.	267
F-34 Descent profiles vs. potential density, ABE-131.	267
F-35 Time series of hydrothermal tracers after pre-processing, ABE-131.	268
F-36 Time series of hydrothermal tracers processed for NBP detection, ABE-131.	269
F-37 Time series of hydrothermal tracers processed for BP detection, ABE-131.	270
F-38 Bird's eye view of anomaly intensity after processing for NBP detection, ABE-131.	271
F-39 Bird's eye view of anomaly intensity after processing for BP detection, ABE-131.	272
F-40 Bird's eye view of ADCP-derived crossflow velocity, ABE-131.	273
F-41 Descent profiles vs. depth, ABE-133.	274
F-42 Descent profiles vs. potential density, ABE-133.	274
F-43 Time series of hydrothermal tracers after pre-processing, ABE-133.	275
F-44 Time series of hydrothermal tracers processed for BP detection, ABE-133.	276

F-45 Bird's eye view of anomaly intensity after processing for BP detection, ABE-133.	277
F-46 Bird's eye view of ADCP-derived crossflow velocity, ABE-133.	278
F-47 Descent profiles vs. depth, ABE-134.	279
F-48 Descent profiles vs. potential density, ABE-134.	279
F-49 Time series of hydrothermal tracers after pre-processing, ABE-134.	280
F-50 Time series of hydrothermal tracers processed for BP detection, ABE- 134.	281
F-51 Bird's eye view of anomaly intensity after processing for BP detection, ABE-134.	282
F-52 Vent prospecting summary data, SMAR.	283
F-53 Closeup of vent prospecting summary data, SMAR.	284
F-54 Descent profiles vs. depth, ABE-150.	285
F-55 Descent profiles vs. potential density, ABE-150.	285
F-56 Time series of hydrothermal tracers after pre-processing, ABE-150.	286
F-57 Time series of hydrothermal tracers processed for NBP detection, ABE- 150.	287
F-58 Time series of hydrothermal tracers processed for BP detection, ABE- 150.	288
F-59 Bird's eye view of anomaly intensity after processing for NBP detec- tion, ABE-150.	289
F-60 Bird's eye view of anomaly intensity after processing for BP detection, ABE-150.	290
F-61 Bird's eye view of ADCP-derived crossflow velocity, ABE-150.	291
F-62 Descent profiles vs. depth, ABE-151.	292
F-63 Descent profiles vs. potential density, ABE-151.	292
F-64 Time series of hydrothermal tracers after pre-processing, ABE-151.	293
F-65 Time series of hydrothermal tracers processed for BP detection, ABE- 151.	294
F-66 Bird's eye view of anomaly intensity after processing for BP detection, ABE-151.	295
F-67 Bird's eye view of ADCP-derived crossflow velocity, ABE-151.	296
F-68 Descent profiles vs. depth, ABE-153.	297

F-69 Descent profiles vs. potential density, ABE-153.	297
F-70 Time series of hydrothermal tracers after pre-processing, ABE-153. . .	298
F-71 Time series of hydrothermal tracers processed for BP detection, ABE-153.	299
F-72 Bird's eye view of anomaly intensity after processing for BP detection, ABE-153.	300
F-73 Bird's eye view of ADCP-derived crossflow velocity, ABE-153.	301
G-1 Exact OG map produced from Phase-2 dive ABE-151.	305
G-2 IP algorithm version of the exact OG map in Fig. G-1.	306
G-3 Standard algorithm version of the exact OG map in Fig. G-1.	307
G-4 Boxplots of errors in the posteriors computed by the IP and standard OG algorithms for various priors.	309
G-5 Expected number of vent fields from OG maps produced by the exact, IP and standard algorithms for various priors.	310
G-6 OG algorithm efficiency for various priors.	311

List of Tables

3.1	Parameter values used in simulations.	89
4.1	Maximum likelihood estimates of lumped plume spread and characteristic vent field size.	167
5.1	Maximum likelihood estimates of characteristic source size b_0 for each dive that included interception of BP water.	183
F.1	ABE dives studied in this thesis.	238

List of Acronyms

ABE	Autonomous Benthic Explorer
ADCP	acoustic Doppler current profiler
AUV	autonomous underwater vehicle
BP	buoyant plume
CID	Conditional Independence of Detections
CIM	conditional independence of measurements
CPT	chemical plume tracing
CTD	conductivity; temperature; depth
DVL	Doppler velocity log
ELSC	Eastern Lau Spreading Center
IP	Independence of Posteriors
LBL	long baseline
OBS	optical backscatter
OG	occupancy grid
NBP	non-buoyant plume
RANSAC	RANdom SAmple Consensus

ROV	remotely operated vehicle
SMAR	Southern Mid-Atlantic Ridge
SOG	speed-over-ground
STW	speed-through-water
VOG	velocity-over-ground
VTW	velocity-through-water
VVA	vertical velocity anomaly

Chapter 1

Introduction

Hydrothermal vents are deep ocean phenomena that likely play important roles in the global heat budget and various global chemical budgets, support unique ecosystems, and offer clues about the geological processes that regulate the formation of oceanic crust [40, 49]. Since the first discovery of a vent field in 1979 [19], enormous scientific attention has been directed toward their study. Nevertheless, Baker and German [5] conclude in a recent study that only 20% of the Earth's 50,000 km to 60,000 km of mid-ocean spreading axes has at most cursorily been surveyed for the presence of hydrothermal venting, and only 10% studied thoroughly enough to identify the locations of individual vent fields on the seafloor.

Seafloor hydrothermal vents emit vast plumes of effluent that can be detected kilometers away from the source vent field itself. While these plumes are interesting in their own right, they also enable what would otherwise be an impossibly small target to be located, though the process is far from trivial. Inferring vent field location from hydrothermal plume data is known as hydrothermal vent prospecting [4]. The value of employing an autonomous underwater vehicle (AUV) as a complementary tool within the arsenal of existing sensors and methods for hydrothermal vent prospecting has now been demonstrated on several successful expeditions in both the Pacific and Atlantic Oceans [39, 65]. However, one aspect of AUVs that was not wholly capitalized upon during these expeditions is their capacity for autonomous decision-making.

The original goal of this work was to enhance the scientific yield from AUV-based hydrothermal vent localization and characterization missions by transferring much of the onus of data processing and survey design onto the vehicle itself. In this respect,

I claim some success. Though hydrothermal vent prospecting remains the primary motivator behind this work, I have striven to abstract the results and methodologies developed as much as possible from that specific application. By focusing on the development of generic frameworks and tools, the methods developed herein have broader application to robotic chemical plume source localization¹ in general, and contain insights into the probabilistic robotic mapping algorithm known as occupancy grid (OG) mapping as well as a novel application for this probabilistic robotics workhorse.

1.1 Background

The spatial relationship between the detectable manifestation of a chemical plume source, the plume itself, and the location of its source is inherently uncertain:

- turbulent processes result in the random motion of plume effluent and spreading of the plume;
- background currents advect the plume and cannot be measured exactly;
- plume structure may depend on processes occupying a wide range of spatial and temporal scales;
- plumes from several sources may interact;
- there may be uncertainty in whether a chemical plume has in fact been detected;
- measured concentration may be only weakly dependent on measurement location.

Of course, plume spread due to turbulence and background advection are responsible for making plumes useful for localizing sources in the first place by producing a relatively large spatial signature compared to the size of the source itself. The relative importance of these factors and the others listed above depends on the nature of

¹The more common term in the literature is chemical plume tracing (CPT); however, I will use the term “chemical plume source localization” preferentially to CPT throughout most of this work because not all methods for finding the sources of plumes actually require following (tracing) the plume itself.

the source, its environment, and the characteristics of the effluent composing the plume itself. Many biologically important plumes, for instance the pheromone plume emitted by a female moth seeking a mate, occupy short time scales and contain a specific, passive tracer. Geophysical plumes, for instance the deep-sea hydrothermal plumes that are the principal application area of this work, often occupy much larger temporal and spatial scales, may include buoyancy effects, and usually contain a variety of tracers rather than one specific chemical.

Potential robotic applications for chemical plume tracing are many and varied. They include pollution and environmental monitoring, chemical plant safety, search and rescue, anti-terrorism, narcotics control, and explosive ordinance removal including demining, and hydrothermal vent prospecting as explored in this thesis. Potential algorithmic solutions to the chemical plume tracing problem range from gradient ascent, to biologically inspired algorithms, to strategies that rely on building maps to estimate source location, with some recent work also exploring multi-agent cooperative approaches. In all cases, successful search requires an algorithm designed to capitalize on the scales present in plume structure that carry information about source location.

1.1.1 Examples from the Natural World: Moths, Starfish, Lobsters and Crabs

Examples of olfactory-based localization of odor sources abound in the natural world. Lobsters, crabs and moths in particular all have extensive literatures devoted to describing the aspects of their behavior that result in the robust and efficient tracking of odor sources that these animals routinely execute (for recent overviews, see [122, 126]).

Even in environments with comparatively high Reynolds numbers, dominated by turbulent flow, sufficiently long time scales combined with a means to sense instantaneous gradient may make classical (mean-gradient) chemotaxis a viable strategy. Given enough time, the average concentration of a turbulent plume does provide gradient information as to the location of its source. This strategy may be the method by which starfish find the plumes of altered water emitted by their bivalve prey [21, 24].

The pheromone tracking ability of male moths of various species is well known (for

an extremely condensed historical overview of the literature, see Ludlow [75]). The environmental conditions in which male moths seek out females emitting pheromone are characterized by wind-driven turbulent advection that produces a plume both spatially and temporally variable on timescales relevant to those imposed by the moth’s flight. In this case, behavior tuned to advective timescales makes sense because wind direction provides the most direct clue about source location relative to the searcher.

The principal complication in tracking odor plumes at these scales is the intermittency of the pheromone signal caused by small-scale turbulence in the mean flow [69, 91, 122, 124]. Meso-scale eddies on the order of meters result in meandering plumes [90] such that maintaining intermittent contact with the plume becomes a critical component of the tracking strategies employed by moths [54]. Though there is some debate as to the underlying mechanism, a plume-tracking moth’s behavior can be segmented generally into surging and casting [54]. In essence, male moths “surge” upwind at some angle to the wind upon encountering pheromone above some threshold concentration, but switch to cross-wind lateral excursions (“casting”) after sufficient time without further stimulus. The result is a zigzagging pattern that gradually progresses upwind toward the source. Both simulations and experiment convincingly point to the utility of zigzagging across an odor plume as a strategy well suited to maintaining intermittent contact with the plume [9, 69, 122]. This strategy is known as odor-gated anemotaxis, since it relies on sensing wind direction, but seems to require only binary chemical detection.

Relatively large and slow-moving aquatic creatures such as blue crabs and lobsters likely use a combination of chemical cues and up-current motion (rheotaxis) to locate odor sources emitted by carrion, prey, or other individuals [122, 124]. Blue crabs and lobsters forage in a turbulent boundary layer over the coastal and estuarine sea floor. Chemical plumes in these habitats are thus turbulent themselves and, like atmospheric odor plumes, consist of discrete propagating packets or filaments separated by non-odor-laden fluid [34]. Due to the greater density of water, however, the relevant scales of turbulence are about an order of magnitude smaller than in air [2]. Though lobsters and crabs move more slowly than moths, mean flow velocities encountered in their habitats are typically lower than in air, and they too must react to chemical stimulus on the time scales associated with the small scale eddies that cause the intermittent

nature of the chemical signal [41,122,124,138]. The influence of larger eddies, at scales sufficient to cause the plume to meander, is relatively unexplored due to the limited sizes of laboratory flumes [138], though it is likely that search behaviors exhibited by crabs and lobsters must also be robust to occasional loss of intermittent contact with the plume [124]. The relatively large size of these arthropods is important because it is comparable to the typical near field widths of odor plumes found in their environment. These animals are likely to have evolved to exploit the additional spatial information available in the plume to improve the efficiency of their search [2,3,122], for instance to remain closer to the centerline of the plume [125].

1.1.2 Robotic Plume Source Localization

Robotic approaches to chemical plume source localization can be classified into two broad categories: (1) biomimetic strategies that seek to emulate the remarkable feats of plume tracing in the animal and microbial worlds; (2) model-based strategies that rely on analytical models for plume evolution to invert records of concentration and flow measurements for source location. Multi-agent methods suited to either of these strategies have also been proposed.

Biomimetic Approaches

Grasso [43] states, “Biomimetics operates on the premise that animal behavior serves as an existence proof of a solution.” Numerous authors have attempted to implement on robots biomimetic solutions to the problem of tracing a chemical plume to its source [18,32,43,52,62,92]. While many of these attempts proved successful in the sense that the robot was able, at least some of the time, to locate the odor source, the results have generally failed to match the performance of the creatures whose behaviors researchers were seeking to emulate [41]. This is not altogether surprising considering that the biological algorithms remain incompletely characterized (see the preceding section), and that current olfactory sensor technology still does not rival its natural counterpart [52].

Farrell et al. [29] report a successful plume tracing algorithm inspired by moths and implemented on a REMUS AUV [1]. Their Chemical Plume Tracing (CPT) algorithm consisted of six behaviors switched by chemical detection events and timeouts. The

several versions of the RoboLobster [18,42,43] are an attempt to emulate the behavior of chemical plume tracking lobsters, specifically to investigate the role that spatial information available in the plume plays in the algorithm. As of 2002, live lobsters easily outperformed the RoboLobster, but the work showed conclusively that spatial information can improve the efficiency of biomimetic CPT.

Within the terrestrial robotics community, Kuwana et al. [60–63] and Nagasawa et al. [92,93] circumvented the technological limitations imposed by the sensitivity and response time of artificial olfactory sensors by attaching live antennae from a silkworm moth *Bombyx mori* to a series of silkworm moth mimics called PheGMots (Pheromone Guided Mobile Robots). Two European groups at the University of Tübingen in Germany [70–72,123] and Örebro University in Sweden [23,28] are working to design robots (“electronic watchmen” [72]) capable of localizing chemical sources in indoor environments without strong and persistent mean flows. Although not explicitly biomimetic, the algorithms employed are either strictly reactive or driven by trained neural networks.

Model-based Approaches

A different approach to robotic plume source localization pursued by some authors requires estimating or assuming the parameters of some model of plume formation while concurrently inverting that model for source location. Unlike biomimetic methods, navigation with bounded uncertainty is requisite because some spatial representation of acquired data must be maintained. Ultimately, the results are subject to how well the model represents the environment. To my knowledge, all previously reported results apply only to single source scenarios.

Ishida et al. [51] developed a terrestrial system that estimates the parameters (including source location) of a time-averaged model of plume dispersal in a uniform advective field. Their robot was able to successfully locate an ethanol source a few meters away from its starting location in several minutes. The slow convergence time and limited range are a consequence of the time required for average concentrations in the actual plume to converge to those predicted by the model. Christopoulos and Roumeliotis [15] describe an algorithm for optimally adapting robot trajectory to estimate the parameters of an advection/diffusion model of plume evolution and present simulation results.

Farrell et al. [32] developed a plume mapping and source localization approach based on hidden Markov methods to concurrently estimate the likelihood of odor detection versus position, the likelihood of source location versus position, the most likely path taken by the odor to a given location, and the path between two points most likely to result in odor detection. This approach is suited to strongly advective environments when the width of the plume is small relative to the search area.

Pang [101,102] has developed a Bayesian method for updating the probability that discrete cells on a grid contain the source of a chemical plume. His algorithm relates to the approach pursued here and is discussed in more detail in Ch. 2. Advances reported in this work include: applicability to multiple-source domains, and applicability to a class of sensor model rather than to a specific type of chemical plume.

Multi-Agent Approaches

Several authors have developed multi-agent plume source localization algorithms. The basic idea is that search times can be reduced by sharing information across a distributed group of robots. “Biologically-inspired” methods [33,44] rely on each robot to execute its own plume search algorithm with communication between agents used to direct the swarm toward robots having the greatest success. Alternate methods use robots as nodes in a distributed sensing network to estimate the parameters of a plume model (including source location) [16] or to instantaneously compute spatial gradients toward source location [136].

1.1.3 Hydrothermal Vent Prospecting

Although the plumes of altered water emitted by hydrothermal vents are often readily detectable with standard in situ sensors to within kilometers of seafloor vent sites [5-7], the physical characteristics of hydrothermal plumes make pinpointing vent sites on the seafloor time-consuming and challenging. Ship-based conductivity; temperature; depth (CTD) vertically oscillating tows or “tow-yos” are the primary means of determining the presence of hydrothermal venting [6] with towed arrays of temperature and optical backscatter probes also now widely employed [7]. One pass along the axis of a mid-ocean ridge is generally sufficient to detect the presence of large-scale (100 + MW) venting and localize its source to within a few kilometers,

perhaps less if high resolution bathymetry of the search area is available. Following initial detection, further CTD tows may be conducted to map the plume and infer the source location. This process is made challenging by the fact that the plume is changing significantly on scales of a few hours due to tidal currents, and concurrently represents the integrated output of several tidal cycles worth of discharge (Fig. 4-2). At some point, near-bottom assets must be deployed. The degree to which the area is constrained, over which near-bottom assets requiring surface supervision must search before locating the vent field, significantly impacts the scientific return of the cruise.

This is particularly true when the assets available require dedicated use of the ship from which they are deployed. Towed, tethered, and occupied assets all require the attention of a dedicated ship; however, towed assets remain indispensable for establishing initial contact with a hydrothermal plume, and tethered or occupied vehicles remain the only option for tasks involving sampling, detailed inspection, and manipulation of objects or instruments on the seafloor. The intermediate task, that of localizing and initial characterization of a hydrothermal source once within the plume, is one to which AUVs are uniquely suited and for which we have had considerable success [39,65]. This portion of the localization process is essentially one of mapping the plume, with multiple sensing modalities each offering their own insight into a plume’s probable source.

Several approaches to chemical source localization specific to hydrothermal plumes have been proposed. Veirs et al. [121] propose a method whereby CTD-based detections of density inversions are back-propagated to their probable sources using records of current velocity between the time of detection and a time in the past based on the theoretical maximum equilibration time of hydrothermal plumes. This method successfully identified the locations of several known vent fields and suggested locations that might contain undiscovered sites. Lavelle et al. [67] describe an inverse calculation whereby temperature and current velocity records from a stationary mooring were inverted for source locations and associated buoyancy flux on a discrete grid. A so-called “puff” model of plume dispersal, along with a four day record of currents and temperature were sufficient to resolve locations for all known sources in the region. Finally, Burian et al. [11] describe several gradient ascent methods specifically intended for implementation on an AUV and robust to local concentration maxima.

1.2 Approach

This thesis develops tools to enable chemical plume source localization via the autonomous execution of nested surveys. The essential characteristics of a nested survey are that each successive stage be smaller in extent and provide higher resolution information about the feature of interest than the previous stage, thereby enabling the surveying vehicle to “home-in” on the target(s). Nested surveys are a conceptually intuitive methodology for locating small features in a large environment when those features have manifestations in the environment larger than themselves (e.g. the plume emitted by a chemical source). Figure 1-1 summarizes the components of approach and Figure 4-4 shows an example of a set of nested AUV surveys employed to localize the source vent field of a hydrothermal plume. Higher resolution data associated with source location was attained on each successive survey by both descending in the water column toward the seafloor and by tightening trackline spacing.

A strategy based on nested surveys is just one of many search methods potentially applicable to the chemical plume source localization problem. Figure 1-2 places the nested survey approach in context with the alternate strategies discussed in the preceding section. All fall into one of two broad groups differentiated by the strength of the coupling between vehicle trajectory and sensory input. Strategies with weak coupling (map-based) rely on abstracted representations of sensor data in terms of probable source location to adapt vehicle trajectory, and often only after long periods of no adaptation. Strategies with strong coupling (behavior-based or reactive) modulate vehicle motion directly in response to sensor input, usually over short timescales. Among the map-based strategies, nested surveys are intermediate between completely preplanned surveys with no data-driven component and fully adaptive trajectory generation wherein trajectory is continuously modulated.

1.2.1 Nested Surveys for Chemical Plume Source Localization

I chose to pursue automation of a nested survey approach to chemical plume source localization in this work for the following reasons:

- applicability to multi-source domains;

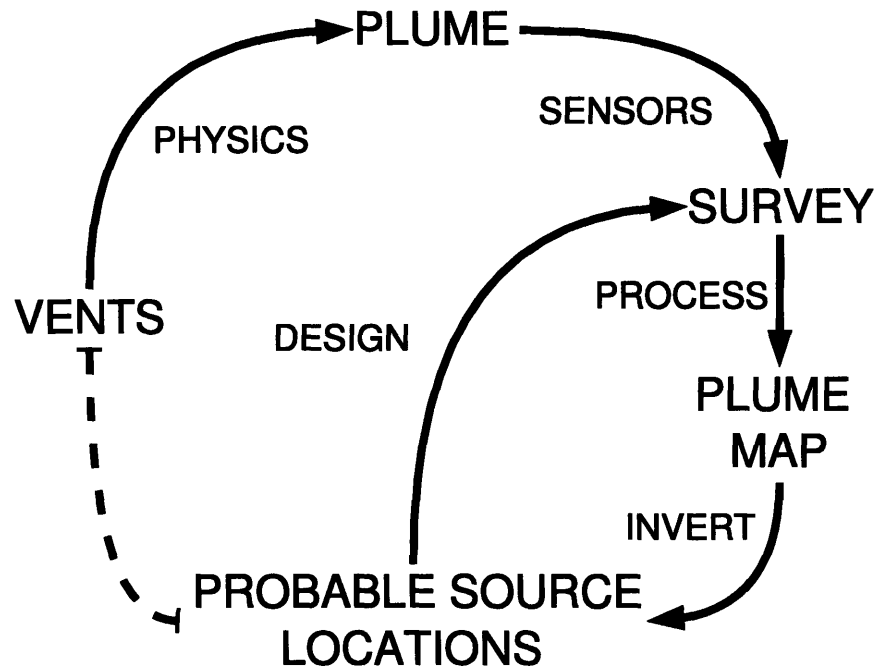


Figure 1-1: The nested survey approach to chemical plume source localization. Sources release plumes into the environment which are mapped through a surveying vehicle's sensors and its trajectory onto a timeseries of measured concentration. Processing yields a map of the plumes themselves which should emphasize scales that contain information about source location. Further processing yields a map of probable source locations and of locations unlikely to contain sources. This information drives the design of the subsequent survey stage and the process repeats. With the completion of each successive high-resolution survey, uncertainty in source locations should diminish, culminating in groundtruthed source locations.

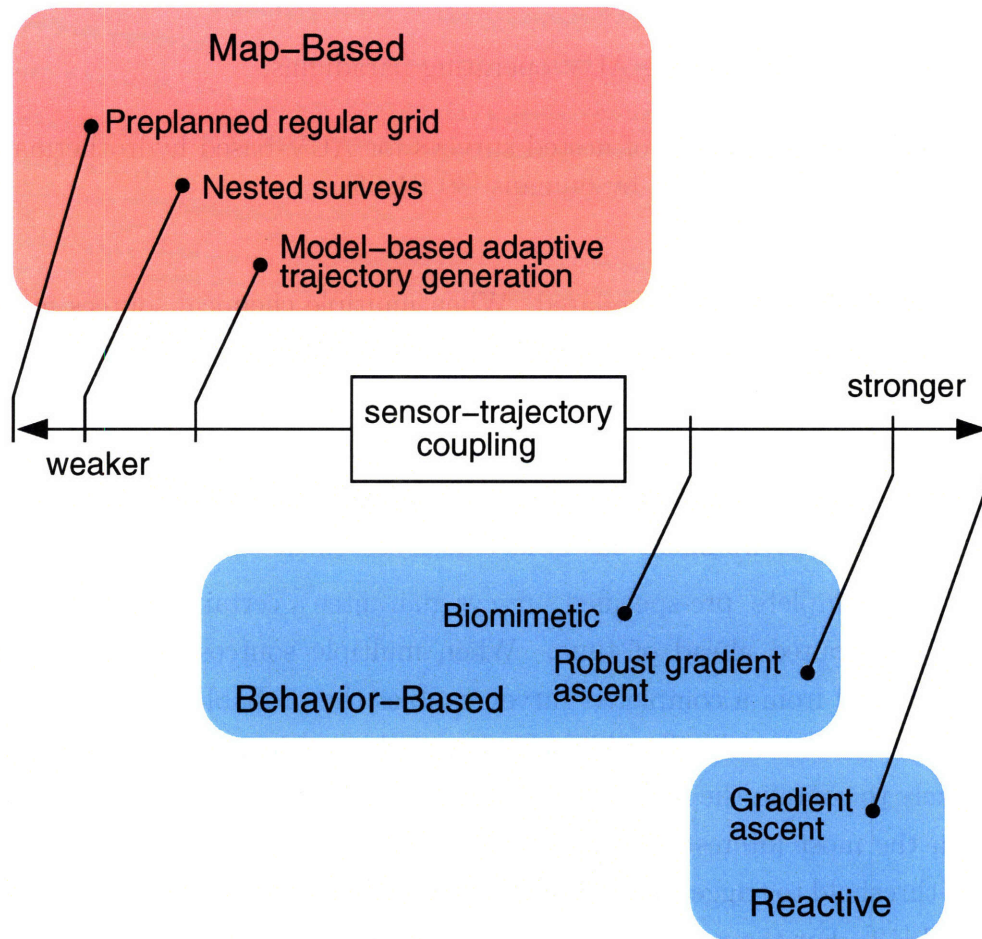


Figure 1-2: A spectrum of chemical plume source localization methodologies differentiated by the strength of coupling between vehicle trajectory and sensory input. A strategy based on nested surveys as pursued in this work alters vehicle trajectory only after completion of a survey stage. All data from the previous stage can be used to drive design of the subsequent stage, if encapsulated in a map of likely source locations. The tradeoff is attainable efficiency. A behavior-based or reactive strategy well-matched to type of plume being traced can potentially home-in on a source much faster; however, these strategies only apply to single source domains, although ad hoc modifications [31] have been employed to overcome this limitation.

- guaranteed coverage of the prescribed survey area;
- robustness to low-value targets and false alarms;
- compatibility with existing AUV operating paradigms;
- the demonstrated success of nested surveys for AUV-based hydrothermal vent prospecting when directed by humans [39, 64, 65].

The first three bullets above are related. When multiple chemical sources might be present in the survey domain, it is often important to find them all. For example, a navy may wish to be confident a harbor is completely free of mines. Tight coupling of vehicle trajectory generation to sensory input provides little guidance on how to proceed once a single mine has been found. It is difficult in such contexts to ascertain what portion of the environment has in fact been searched.

In contrast, complete, pre-specified surveys guarantee a certain degree of coverage that can be computed ahead of time. When multiple sources might be present, the map produced from a completed survey enables the probable number of sources present to be assessed and will guide selection of the portions of the search domain that are likely to contain the highest value targets. This approach permits the vehicle to focus on the most interesting aspects of the data, rather than relying on careful design of a threshold to trigger pursuit of all potentially interesting features as they are encountered [12]. For the problem of detecting and locating persistent hydrothermal vents on the seafloor, nested surveys may require the vehicle to cover more ground than a triggered approach would. However, the vehicle is less likely to spend time exploring false alarms or low-value targets.

The last bullet refers to the fact that typical AUV operations involve flying regular tracklines in a grid pattern so as to uniformly sample a prescribed area. While regular tracklines may not be necessary for plume survey in general, they are necessary, for instance, when collecting acoustic bathymetry or optical imagery for photomosaicking. Nested surveys do not require alteration of vehicle trajectory, meaning these other aspects of a survey can continue unimpeded, perhaps dictating some aspect of the nested surveys themselves.

Successful execution of a nested survey strategy relies on the creation of maps at each stage of the survey to guide the design of the follow-on stage. The key survey design elements that must be drawn from each map are good choices for the location and extent of the subsequent survey stage. Survey resolution might also be guided by the map or alternately fixed by the known characteristics of the feature of interest or by constraints imposed by other sensing modalities. A rudimentary map might simply comprise the locations at which the plume was detected, and perhaps the concentrations of those detections (which may or may not contain significant information about source location depending on the nature of the plume). Of more direct utility are maps that encapsulate the likely locations of the sources themselves. That requires processing raw sensor data through a model of plume evolution, that is, inversion of the plume for source location.

As the relationship between plume location and concentration is inherently uncertain, I conjecture that constructing maps of source locations that include an assessment of the uncertainty in these locations will facilitate autonomous survey design. Such an approach falls within the rubric of *probabilistic robotics*, the central goal of which is to represent information in the form of probability densities [115,118]. Maps that are suitable for autonomous interpretation in terms of follow-on stage design must offer clear indications of high-value regions and the survey extent required to cover them. The probabilistic map-making methodology pursued in this thesis is known in the robotics literature as occupancy grid (OG) mapping.

1.2.2 Occupancy Grid Mapping and its Application to Chemical Plume Source Localization

OG maps discretize the environment into a collection of cells arranged in a regular pattern wherein each cell has a binary state, either occupied or empty. OG mapping algorithms then generate estimates of the posterior probabilities of occupancy and emptiness of each grid cell. Most occupancy grid update rules recursively incorporate new measurements thereby enabling real time operation. A recursive update rule is not strictly necessary for a survey strategy that calls for completion of predetermined survey patterns prior to map analysis. Nevertheless, this property remains attractive because it broadens applicability to hybrid strategies that might not require comple-

tion of a full survey stage before adapting vehicle trajectory.

OG mapping was originally developed at the Carnegie-Mellon University Mobile Robot Laboratory as a means of incorporating imprecise information about the range from a mobile robot to nearby obstacles acquired from low-cost wide-angle sonar range finders into a detailed map of empty and occupied space [25,85,87].² The same representation is adaptable to chemical plume source localization by redefining occupancy to indicate a cell contains an active chemical plume source rather than indicating the presence of an obstacle in the cell. Occupancy grid mapping has traditionally been restricted to the mapping of static environments, though recent extensions [130,131] are alleviating this restriction. Though chemical plumes are dynamic phenomena, a static OG representation suffices if the static locations of immobile sources are sought, rather than the variable locations of their emitted plumes. Scalar observations of plume presence and absence take the place of range measurements in the typical application, with absence providing constraints on where sources are unlikely to lie.

There are two aspects of the OG representation that make it particularly suited to chemical plume source localization. First, because an occupancy grid map explicitly represents empty space with a degree of confidence, the map can be used to assess not only probable source locations but also whether a survey area has been adequately searched, thereby providing guidance on the extent of follow-on survey stages. Second, since occupancy grid maps are not maps of feature locations, they do not require the additional overhead of reliably identifying recognizable features and then differentiating between individual features. This property is important for multiple source scenarios, where the number of sources cannot be known a priori.

1.3 Contributions

This dissertation contains contributions on three fronts: (1) occupancy grid mapping; (2) general chemical plume source localization; and (3) hydrothermal vent prospecting. The following lists the specific contributions in each category.

²These references predate use of the term “occupancy grid” in the published literature. Instead, these earlier references refer variously to “occupancy maps,” and “certainty grids.” The latter reflects the realization that other types of information besides just occupancy can be represented [79]. The term “Occupancy Grid” appears to have been coined by Elfes [26] in the title of his doctoral thesis.

Occupancy Grid Mapping

- The application of occupancy grid mapping to a novel arena–multi-source chemical plume source mapping.
- An articulation of the problems associated with the application of standard Bayesian occupancy grid mapping to environments with few expected occupied cells (low prior probability of occupancy) and stemming from a key independence assumption required by the standard algorithm.
- An exact solution for the state of an occupancy grid when measurements consist of binary detections and non-detections generated by a particular form of forward sensor model.
- A family of novel approximate algorithms applicable to low prior environments based on this exact formulation.

Chemical Plume Source Localization

- An abstracted forward model for binary chemical plume detection that encapsulates the role of multiple sources without reference to the physics of a particular type of plume.
- A plume source location mapping method suitable for use in multi-source environments.

Hydrothermal Vent Prospecting

- A procedure for automatic classification of hydrographic data into the background water and the two main components of a hydrothermal plume–the buoyant and non-buoyant plume–and its application to field data.

- An algorithm for the generation of occupancy grid maps of the seafloor showing locations likely to contain hydrothermal vents and also regions unlikely to contain vents. These maps are shown to be suitable for the automation of nested surveys in support of autonomous hydrothermal vent prospecting.
- A simple model for buoyant hydrothermal plume evolution suitable for use with occupancy grid mapping methods.
- An evaluation of the utility of measuring crossflow velocity on a surveying AUV for constraining the source locations of encountered buoyant plumes.

1.4 Document Structure

This dissertation is divided into two parts intended to separate the more broadly applicable theoretical contributions of this thesis from the specific application of those methods to AUV-based hydrothermal vent prospecting. This chapter provided an introduction to the problem of chemical plume tracing and specifically its application to hydrothermal vent prospecting. Chapter 6 ties the specific contributions of this thesis back into this wider context and suggests productive avenues for further research. The following provides an overview of the intervening chapters.

Readers interested only in the application of AUVs to hydrothermal vent prospecting can begin at Part II, but should familiarize themselves with occupancy grid mapping (Ch. 2, § 2.1) before reading Ch. 5.

Part I: Bayesian Occupancy Grid Mapping with Binary Measurements in Environments with Few Occupied Cells

Chapter 2: Exact Occupancy Grid Mapping for Binary Measurements

This chapter explores the application of occupancy grid mapping to the chemical plume source localization problem, and more generally to environments with few expected occupied cells (low prior probabilities of occupancy). I begin by developing a generic forward model for binary detection of chemical effluent from multiple

sources. Initially I had expected to be able to then apply standard occupancy grid mapping methods without modification. This chapter reveals why a key independence assumption required by the standard algorithm fails to produce accurate maps in this application. However, special properties inherent in the forward model allow for existence of an exact solution computable under certain conditions. In particular, it is shown that non-detections can be incorporated into an occupancy grid map recursively, and without approximation.

Chapter 3: Approximate Algorithms for Low Prior Environments The exact occupancy grid mapping algorithm introduced in Ch. 2 incurs linear cost in map size, the penalty for which is exponential cost in the number of detections registered by the sensor. This exponential scaling renders the exact formulation impracticable. Its utility derives instead from providing the foundation for several novel approximate algorithms introduced in this chapter. Two families of algorithms are developed and their performance studied in simulation relative to the exact algorithm and to other metrics of performance.

Part II: Automated Nested Survey for Hydrothermal Vent Localization

Chapter 4: Hydrothermal Plume Survey by Autonomous Underwater Vehicle This chapter introduces hydrothermal plumes as perceived by a surveying AUV. I explore the hydrographic data obtained by the ABE AUV during several recent hydrothermal plume prospecting expeditions in order to develop methods for the automated detection of hydrothermal effluent, and to develop constraints on seafloor vent location from the locations of those detections above the seafloor along with measurements of ambient crossflows. The modeling efforts in this chapter form the foundation for the application of occupancy grid methods to the same data in Ch. 5.

Chapter 5: Buoyant Hydrothermal Plume Source Localization via Nested Survey This chapter shows how the occupancy grid mapping methods developed in Part I can be applied to mapping the probable locations of hydrothermal vents on the seafloor. These maps are readily machine-interpretable in terms of the parameters of each survey stage within a nested survey. These automatically generated survey

trajectories are found to compare favorably to the actual human-generated surveys planned in the field.

Part I

Bayesian Occupancy Grid Mapping with Binary Measurements in Environments with Few Occupied Cells

Chapter 2

Exact Occupancy Grid Mapping for Binary Measurements

This chapter presents the derivation of an exact occupancy grid mapping algorithm under the dual restrictions that

- the mapping sensor report only binary detections and non-detections,
- and that the sensor can be modeled with a particular invertible form of forward model to be introduced.

Nominally the computational cost of exactly computing the marginal posteriors would scale exponentially with map size and is consequently unfeasible [118]. This chapter reports an exact algorithm with linear cost in map size, the penalty for which is exponential cost in the number of detections registered by the sensor. While the number of detections expected in an environment with few occupied cells may be quite modest, this exponential scaling renders the algorithm ultimately impracticable. Its utility derives instead from the insights it offers into the nature of occupancy grid mapping in a low-prior environment and by providing the foundation for the several novel approximate algorithms that form the subject of the next chapter.

Though the application pursued in this thesis is underwater chemical plume tracing, the insights offered into OG mapping are more general, having broader relevance particularly in sparsely occupied environments. If observations are limited to scalar measurements of a chemical tracer, atmospheric plume tracing is an obvious application arena; however, the methodology developed in this and the next chapter is

more broadly applicable to any static mapping problem wherein measurements can be distilled to binary detections and non-detections. The archetypal application of indoor mapping with sonar range-finders can be formulated in these terms; though additional work remains to determine whether the methods developed herein offer any advantages over traditional methods. Robotic mapping efforts in sparsely occupied environments should benefit directly. Such environments are usually the purview of feature-based methodologies; however, a map of feature locations includes no intrinsic mechanism to represent confidently unoccupied space, meaning OG methods could play a complementary role in combined mapping/exploration missions.

This chapter begins with a presentation of the classical Bayesian formulation of the OG mapping algorithm [85], the formulation studied and extended in this and the next chapter. The key assumptions required are described along with a review of the literature devoted to exploring the consequences of these assumptions and their circumvention. The validity of these assumptions is then assessed in the context of the localization of chemical plume sources. This analysis provides the motivations for the extensions to the standard Bayesian approach developed next. For the sake of readability, only key results appear in the main text. Detailed derivations have been relegated to App. B.

2.1 Background: Occupancy Grids

Probabilistic mapping algorithms strive to estimate the posterior probability distribution over the space of all maps \mathbf{m} given the set of all measurements $Z^t = \{z^t, z^{t-1}, \dots, z^1\}$ made of the environment: $p(\mathbf{m} \mid Z^t)$ [117]. Since the real world is continuous, the space of all maps is infinite dimensional and this problem is in general intractable. OG mapping algorithms rely on a discretization of the world into a regular grid of cells, each of which is represented by a binary random variable $\mu_c \in \{0, 1\}$ indicating respectively emptiness or occupancy of that cell. The space of all maps is now finite with a dimensionality of 2^C , where C denotes the number of grid cells that make up the map. Nevertheless, computing the posterior probabilities for all possible 2^C maps is a formidable problem for typical OG maps with thousands to millions of grid cells.

To circumvent this issue, OG methods decompose this high-dimensional problem

into a collection of independent, one-dimensional estimation problems, wherein the state of each cell is estimated independently of the states of the remaining cells in the map. Methods based on Bayesian inference [26,27,85], Dempster-Shafer evidence theory [100], Fuzzy set theory [98,99], and various ad hoc rules, particularly in the early literature, exist to update estimates of the state of each cell in the grid as new measurements arrive.¹ The key concession common to all these methods is the inability to represent dependencies between the states of multiple cells. These dependencies arise because sensors typically observe more than a single grid cell during each measurement [117].

2.1.1 Bayesian OG Mapping

This work is concerned with application and extension of the Bayesian method exclusively. The Bayesian approach seeks to estimate the marginal posterior probability densities $p(\mu_c \mid Z^t)$ for all cells c given prior probability densities $p(\mu_c)$. Since these densities all correspond to binary random variables, they are equivalently determined by the knowledge of the posterior probability of occupancy $P[m_c \mid Z^t]$ or of emptiness $P[\bar{m}_c \mid Z^t] = 1 - P[m_c \mid Z^t]$, where m_c and \bar{m}_c denote $\mu_c = 1$ and $\mu_c = 0$ respectively.

A simple recursive update rule for these posteriors is attained if we regard estimating the states of all cells as a collection of independent binary estimation problems. Begin by applying Bayes Rule to the marginal posterior:

$$\begin{aligned} P[m_c \mid Z^t] &= P[m_c \mid z^t, Z^{t-1}] \\ &= \frac{P[z^t \mid Z^{t-1}, m_c] P[m_c \mid Z^{t-1}]}{P[z^t \mid Z^{t-1}]} . \end{aligned} \quad (2.1)$$

To proceed, a critical assumption is required:

$$p(z^t \mid Z^{t-1}, \mu_c) = p(z^t \mid \mu_c) . \quad (2.2)$$

Eq. (2.2) states that the current measurement is conditionally independent of all previous measurements given knowledge of the state of the single cell c . Sensors that

¹Ribo and Pinz [106] provide an insightful comparison of three modern map update rules in a typical office environment.

simultaneously measure the state of multiple grid cells violate this assumption [117]. A sensor that scans multiple cells to generate a single scalar measurement z^t implies that information about the states of grid cells other than the one being updated could help to predict the outcome of the measurement z^t . There must be information about the state of the grid in Z^{t-1} or else there would be no point in having made these measurements. As a result, Z^{t-1} contains information about the probable outcome of z^t not subsumed by knowledge of μ_c , and conditional independence cannot hold. The ramifications of this particular assumption are many and I will refer to it throughout this work. I will use the acronym CIM to denote Conditional Independence of Measurements, leaving implicit that the conditioning required is the state of single cell being updated.

Accepting the CIM assumption for the moment, (2.1) becomes

$$\begin{aligned} P[m_c | Z^t] &= \frac{1}{P[z^t | Z^{t-1}]} P[z^t | m_c] P[m_c | Z^{t-1}] \\ &= \frac{P[z^t]}{P[z^t | Z^{t-1}]} \frac{P[m_c | z^t]}{P[m_c]} P[m_c | Z^{t-1}] , \end{aligned} \quad (2.3)$$

where the second line follows from the reapplication of Bayes Rule to the second term. An analogous procedure leads to the posterior probability of emptiness:

$$P[\bar{m}_c | Z^t] = \frac{P[z^t]}{P[z^t | Z^{t-1}]} \frac{P[\bar{m}_c | z^t]}{P[\bar{m}_c]} P[\bar{m}_c | Z^{t-1}] . \quad (2.4)$$

The lead terms cancel upon constructing the so-called odds ratio $r_c^t \triangleq P[m_c | Z^t] / P[\bar{m}_c | Z^t]$:

$$r_c^t = \frac{P[m_c | z^t]}{1 - P[m_c | z^t]} \cdot \frac{1 - P[m_c]}{P[m_c]} \cdot r_c^{t-1} , \quad (2.5)$$

where we have made use of the binary nature of μ_c to replace all probabilities of emptiness with one minus the corresponding probability of occupancy.

The form of (2.5) is intuitively appealing. Cells whose occupancy is strongly suggested by the current measurement ($P[m_c | z^t] \approx 1$) have their odds of being occupied raised. Analogously, cells whose vacancy is supported by the current measurement ($P[m_c | z^t] \approx 0$) have their odds of being occupied lowered. In each case, the result

includes weighting by the prior odds and the previous estimate of the odds ratio. Though intuitively satisfying and computationally attractive (recursive), (2.5) relies on the fallacy that measurements represent independent observations of individual cells.

Relationship to Joint Posterior

The true relationship between each single-cell marginal posterior and the joint posterior over the whole map is attained by marginalizing over all maps $\{\mathbf{m} : m_c\}$ wherein cell c is occupied:

$$P[m_c | Z^t] = \sum_{\mathbf{m} : m_c} P[\mathbf{m} | Z^t] . \quad (2.6)$$

A low marginal posterior implies the maps with that cell occupied are unlikely given the data, and conversely, a high marginal posterior implies a cell that is occupied in likely maps. This fact accounts for the utility of the occupancy grid representation; however, the algorithm specified by (2.5) does not compute the true marginal posteriors due to the CIM assumption.²

Recall that the updating algorithm decomposes the problem of estimating the joint posterior into a collection of one-dimensional problems. At best, the joint posterior could be approximated by assuming that the posterior probabilities computed this way represent the true marginal posteriors and furthermore that these marginals are independent:

$$p(\mathbf{m} | Z^t) = \prod_{c=1}^C p(\mu_c | Z^t) . \quad (2.7)$$

Note that (2.7) requires the prior probabilities of occupancy for each cell be independent in order to hold in the base case ($Z^t = \emptyset$):

$$p(\mathbf{m}) = \prod_{s \in \{1, \dots, C\}} p(\mu_s) . \quad (2.8)$$

This assumption along with CIM leads directly to (2.7) since independence of the marginal priors will not be broken by a sensor that observes single cells individually (the only type that satisfies CIM). However, the converse, that independence of the

²In fact, it will be shown that the algorithm specified by (2.5) serendipitously produces the true marginal posteriors when conditioned exclusively on non-detections (§ 2.4).

marginal posteriors implies CIM, is false. I show in § 2.3.2 that for a certain form of binary sensor model to be introduced subsequently, the standard update rule (2.5) applied exclusively to non-detections correctly computes the marginal posteriors, and furthermore these marginal posteriors satisfy (2.7). Under those special conditions, (2.5) can be derived without requiring CIM. Indeed, it will be shown that the CIM assumption remains false. By contradiction, (2.7) does not, therefore, imply CIM.

“Static World” Assumption

Though not strictly required to derive (2.5), a common additional assumption in made occupancy grid mapping is that the world is static [117]:

$$p(z^t \mid Z^{t-1}, \mathbf{m}) = p(z^t \mid \mathbf{m}) . \quad (2.9)$$

This assumption states that measurements carry no information about one another not subsumed by knowledge of the entire static map. It is necessary for the CIM assumption (2.2) to hold and is much weaker. Recent research [131] is alleviating this restriction; however, the methods developed in this work all require (2.9) to hold. Moving objects, by definition not part of the static map, violate this assumption because observations of such objects are generally useful to predict the outcome of future observations. To illustrate the point, consider two subsequent detections of a chemical plume in ambient flow at two different locations. Moving the sensor upwind (against the wind) from the first detection location will present the sensor with unsampled fluid and therefore the outcome of the previous measurement is irrelevant to the present measurement except inasmuch as it contains information about the plume’s source. Of course, the conditioning on knowledge of the entire map of source locations subsumes this information. On the other hand, motion downwind (with the wind) would tend to sample the same water thereby ensuring correlation of the previous and present measurements, thereby violating (2.9).

2.1.2 Map Artifacts

Several authors working with sonar range-finder data have noted artifacts in OG caused by violating the CIM assumption and have proposed solutions.³ Moravec and Cho [86] (see also [74]) propose a method wherein new range data is interpreted relative to current state of the map, termed the “context-sensitive” approach. Berler and Shimony [10] propose constructing a Bayes network to represent the dependencies between dynamically defined regions created by the overlap of successive measurements. In their work, the grid representation is used only to visualize the result, since rather than estimating the probability of cell occupancy, they estimate the probability that an obstacle exists somewhere within each region. Konolige [58] developed a significant extension to the occupancy grid framework called MURIEL (MUltiple Representation, Independent Evidence Log) which keeps track of robot poses that have observed each cell so as to ignore highly correlated measurements of a cell from similar poses rather than treating them as independent. MURIEL also uses the observations of a cell to update the probability that some readings were due to specular reflections.

In more recent work, Thrun [117] gives a particularly insightful discussion of the manifestations of the CIM assumption in typical OG maps. He shows that independently estimating the states of individual grid cells means that apparently conflicting data about the occupancy of individual grid cells will get averaged together into intermediate values despite the existence of map configurations that could explain the data without conflict. These conflicts are manifest around doorways and other features whose openings are wide enough for sonar beams from some vantage points to pass through, but obscure others. He demonstrates a batch processing method that

³Much of the work specific to indoor environments revolves around limiting the propensity of specular reflections to declare large portions of the map as unoccupied. This tends to occur when cells are updated independently because information in previous measurements about the rest of the map cannot be used to assess the possibility that the present measurement represents a spurious range reading [10]. Specular reflections refer to the coherent reflection of impinging waves on a surface. Specular reflections occur when surfaces are smooth on the scale of the wavelength. Multipath returns occur when a signal reflects specularly off multiple surfaces before returning to the receiver. As a result the receiver reports an erroneously long range. When surfaces are sufficiently flat to produce specular reflections, multipath returns become increasingly likely at steep angles of incidence, however, specular returns with favorable geometry usually result in a strong return and a correct reading [68]. Thus, while specular reflections can cause multipath returns, specular reflections are not synonymous with multipath returns; however, some of the occupancy grid mapping literature uses these terms interchangeably.

utilizes expectation maximization (EM) to search the space of complete maps for maps that maximize the likelihood of all measurements. The method has the disadvantage that it is not recursive and consequently more computationally intensive.

Additional discussion of these works appears in the conclusion to this chapter where I discuss the possible applicability of the methods developed subsequently to indoor environments.

2.1.3 Low Prior Environments

In contrast to indoor sonar mapping, the dominant aspect of a Bayesian OG mapping approach applied to the problem of chemical plume source localization is the low prior with which the grid must be initialized. The Bayesian prior encodes belief in the likelihood of occupied cells before any measurements have been made. Typical occupancy grid mapping applications assume a prior probability of occupancy of $P[m_c] \in [0.2, 0.5]$, in effect assuming somewhere near one half of the environment is expected to be occupied by obstacles [117]. In Part II of this thesis, the OG mapping framework is applied to example data sets from hydrothermal sites in the deep sea on grids with $O(10^4)$ cells, typically only a few of which contain active vents. Depending on grid cell size, priors of $P[m_c] \in [10^{-5}, 10^{-2}]$ are appropriate in such settings.

Unfortunately, the adverse effects of the CIM assumption tend to be exacerbated in a low prior environment. To see why this might be the case, consider the scenario depicted in Fig. 2-1 in which two successive detections of a chemical plume are registered by a robot. The first detection supports the presence of at least one source within the upwind cone denoted C_1 . Considered independently, the second detection supports the presence of at least one source within the upwind cone denoted C_2 . Considered together, several possibilities emerge, two of which are shown schematically in Fig. 2-2. Clearly a very likely possibility is that there is a source within the intersection denoted $C_1 \cap C_2$. Without a source in $C_1 \cap C_2$ at least two sources are required to account for both detections, one each in the portions of C_1 and C_2 not part of their intersection. A small number of expected sources strongly favors the possibility that minimizes the number of sources required to explain the data, so the first explanation is probably correct.

Now consider the same scenario in the context of updating the marginal posteriors

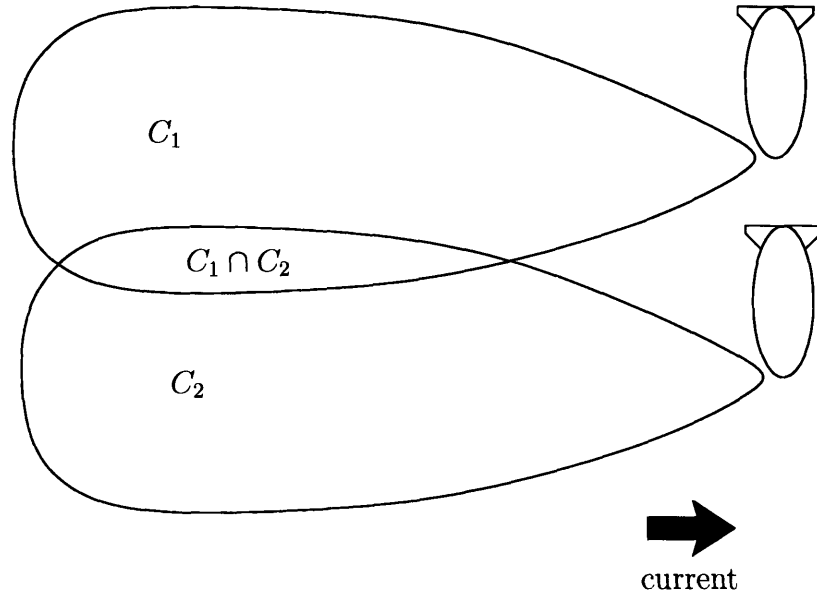


Figure 2-1: Schematic depiction of two detections of a chemical plume made by an AUV showing the upwind regions in which the source or sources of the detected chemical are likely to lie. The current is setting to the right.

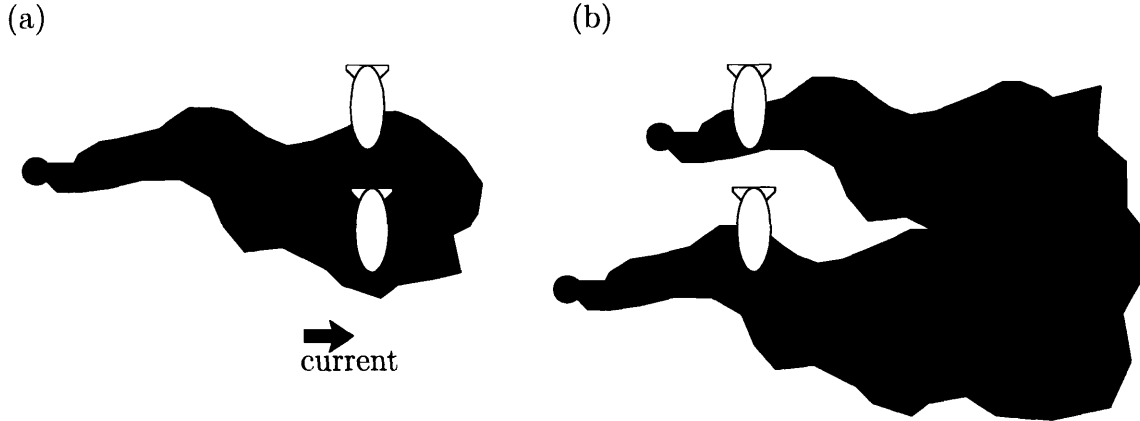


Figure 2-2: Candidate explanations for two plume detections made by a surveying AUV: (a) one source, vehicle relatively far downwind; (b) two sources, vehicle relative close to both.

within an OG map subject to the CIM assumption. Cells that fall within C_1 will have their posteriors increased following the first detection, as will cells that fall within C_2 following the second. Consequently, the posteriors of cells within the intersection $C_1 \cap C_2$ will have increased twice. This is consistent with the notion that maps having a source within the intersection $C_1 \cap C_2$ are now more likely.⁴ However, the attendant

⁴In fact, the posteriors within the intersection $C_1 \cap C_2$ will have increase too much. Because the

effect that maps having occupied cells within the portions of C_1 and C_2 not part of the intersection are now less likely than after either detection considered alone⁵ cannot be captured by an algorithm that estimates of the state of each cell independently, by ignoring the information present in previous measurements about the rest of the map. There is no mechanism in such an algorithm to seek maps consistent with all measurements, and consequently no mechanism to ensure the marginal posteriors reflect the likelihood of occupancy across all maps. This is true regardless of the prior probability of occupancy in each cell; however, a low prior exacerbates the errors incurred because maps with few numbers of occupied cells should be strongly favored.

I emphasize that all OG methods, whether Bayesian or otherwise, will fail to produce exact results in the scenario depicted by Fig. 2-1. This is because all OG methods estimate the state of each cell independently, interpreting each measurement without regard to explanations for that measurement external to the specific cell being updated. An environment with few occupied cells exacerbates the ramifications of this property.

Previous Work

To my knowledge, no previous work directly addresses the consequences of the CIM assumption in Bayesian OG maps initialized with a low prior. Though distinct from OG methods, Pang [101] includes the description of a Bayesian methodology in his doctoral thesis for learning the state of a binary random field of possible locations for a single source from binary detections and non-detections of chemical effluent along with a record of advective currents. His method is developed specifically for short time-scale plumes amenable to modelling [8, 30] as a collections of independently diffusing “filaments” executing biased random walks due to turbulent motions with superimposed advection. The key probabilistic quantity attained from this perspective on plume evolution is, in the notation of [101], $S_{ij}(t_l, t_k)$ which denotes the

detections overlap, they are correlated. Though both detections support the presence of a cell in the intersection $C_1 \cap C_2$, they are not independent observations, and the sum total of their information content is less than if they were.

⁵Because a source in the intersection $C_1 \cap C_2$ explains both detections, the measurements carry correspondingly less information about the remaining portions of C_1 and C_2 then when considered independently.

probability that a source in cell i released a single chemical filament at time t_l given that the filament was present in cell j at time t_k . For cells i that lie far from an upwind trajectory terminating at cell j , $S_{ij}(t_l, t_k) \rightarrow 0$. Because only a single source is assumed, the goal of his method is distinct from that of OG mapping which seeks the marginal posterior probabilities of occupancy without constraining the total number of occupied cells, nevertheless, his grid-based representation of possible source locations provided some of the early inspiration for pursuing OG methods, and his modelling efforts are amenable to incorporation into the OG framework developed later in this chapter.

Pang’s assumption of a single source in the search domain provides a powerful constraint. This constraint ultimately leads to an update rule for a continuously releasing source that enables cells in portions of the grid not actually observed at a given time to nevertheless have their posteriors updated. Consider that the detection of plume effluent at one location means all potential source locations not upwind of the detection are unlikely to contain the source. Therefore, the assumption of a single source has enabled information about all cells in the map to be gleaned from a single measurement, regardless of which cells lie upwind of the measurement location. A similar argument applies to measurements where no effluent is observed. Simulation results presented in [101] demonstrate the applicability of the method to single-source scenarios; however, its extension to multiple-source scenarios is not straightforward because key probabilistic quantities are formulated assuming the existence of only a single source. Appendix E contains a thorough analysis.

The remaining sections of this chapter are devoted to an exact formulation of the marginal posteriors for a specific form of sensor model. In the next chapter, we turn to novel approximate algorithms that relax or eliminate the CIM assumption in favor of alternate assumptions.

2.2 Forward and Inverse Sensor Models

The work herein is applicable to a particular form of invertible forward sensor model. Forward sensor models are distinct from the typical inverse modelling required to interpret sensor measurements in standard occupancy grid mapping. To paraphrase Thrun [117]:

Forward models predict measurements made from within a known map, that is, they model the physics of the sensor. A probabilistic forward model has the form $p(z^t \mid \mathbf{m})$. This probability density function specifies a distribution over measurements z^t given a map \mathbf{m} .

Inverse models attempt to reason from a sensor measurement to its causes within an unknown map. The form of probabilistic inverse model required by standard occupancy grid mapping methods is $p(\mu_c \mid z^t)$, which specifies the marginal probability of occupancy ($\mu_c = m_c$) and emptiness ($\mu_c = \bar{m}_c$) for a single cell c given the measurement z^t .

Because forward models attempt to capture the the physics of the sensor they are arguably more natural. Arbitrarily complex physical phenomena may be included, hence they are also potentially more accurate.

In principle, Bayes Rule applied to any forward sensor model will generate a corresponding inverse model $p(\mathbf{m} \mid z^t)$. The inverse model required by the standard OG algorithm could then be attained by marginalization over all maps:

$$P[m_c \mid z^t] = \frac{\sum_{\mathbf{m}: m_c} p(z^t \mid \mathbf{m}) p(\mathbf{m})}{\sum_{\mathbf{m}} p(z^t \mid \mathbf{m}) p(\mathbf{m})} . \quad (2.10)$$

Marginalization is a form of averaging. The marginal inverse model $P[m_c \mid z^t]$ is the normalized sum of the probabilities of all maps in which cell c is occupied. The large space of all maps (2^C) renders such marginalization by direct summation intractable for most models [118]; however, the form of binary forward sensor model introduced next admits the analytical evaluation of these sums.

2.3 An Invertible Forward Model for Binary Measurements

Let $\delta^t \in \{0, 1\}$ denote a binary-valued random measurement at time t , with z^t reserved for generic measurements. Further, let d^t denote the event of a detection ($\delta^t = 1$) and \bar{d}^t the event of a non-detection ($\delta^t = 0$). The two probabilities $P[d^t \mid \mathbf{m}]$ and

$P[\bar{d}^t | \mathbf{m}] = 1 - P[d^t | \mathbf{m}]$ then completely specify a forward sensor model. The work described herein is applicable to a restricted class of binary forward sensor models that satisfy

$$P[d^t | \mathbf{m}] = 1 - (1 - P_F^t) \prod_{s \in \mathcal{S}} (1 - P_s^t) \quad (2.11)$$

where

$P_c^t \triangleq P[\text{Sufficient signal from occupied cell } c \text{ arrives to trigger a detection at time } t.]$

$P_F^t \triangleq P[\text{A false alarm occurs at time } t.]$

$\mathcal{S} \triangleq \{s : m_s, s \in \{1, \dots, C\}\}$ the set of occupied cell indices in \mathbf{m} .

The subscripts s and c both denote cell index, with s used within products and c externally. False alarms are independent of legitimate detections so that

$$\begin{aligned} P[\bar{d}^t | \mathcal{S} = \emptyset] &= (1 - P_F^t) \\ P[\bar{d}^t | \mathcal{S} = c] &= (1 - P_F^t)(1 - P_c^t) . \end{aligned} \quad (2.12)$$

In words, the first relation above states that the probability of a non-detection given an empty map is equal to the probability that no false alarm occurred. The second states that the probability of a non-detection given a map with only a single occupied cell is equal to the probability that a false alarm did not occur and that the single occupied cell did not present sufficient signal to the detector to trigger a detection.

If all occupied cells of a map have independent probabilities of triggering a detection, then it follows that

$$P[\bar{d}^t | \mathbf{m}] = (1 - P_F^t) \prod_{s \in \mathcal{S}} (1 - P_s^t) . \quad (2.13)$$

Eq. (2.11) then follows trivially. Eq. (2.13) states that the probability of a non-detection is the probability that no false alarm occurred times the probabilities that no occupied cell presented sufficient signal to the detector to trigger a detection.

This model admits the possibility that multiple cells could simultaneously present sufficient signal to the detector to cause a detection; however it disallows occupied cells from reinforcing one another's signals to trigger a detection when the signal present at the detector from each individual occupied cell is insufficient on its own. This

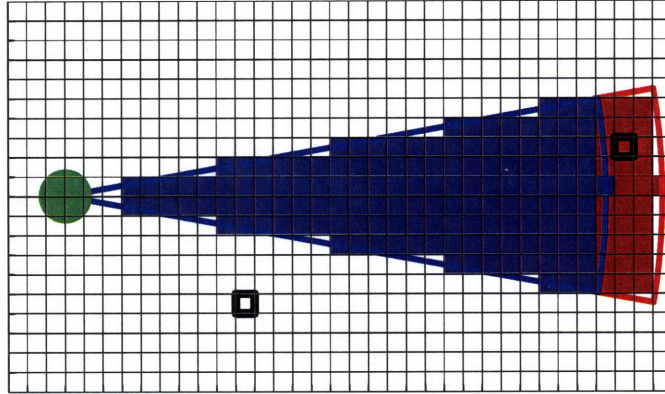


Figure 2-3: Schematic representation of a sonar range-finder reading decomposed into a binary detection and non-detection. Instead of returning a single range after each ping, the model decomposes each ping into two measurements: a non-detection that observes cells in a sector with a radius less than the would be reported range, and a detection that observes cells in an arc centered at the reported range. Similar decompositions have been employed by others [10]. Occupied cells are outlined in black. The non-detection region is highlighted in blue, the detection region in red.

restriction makes it impossible to infer the number of occupied cells in an observation region following a detection, except that at least one must be occupied, or else a false alarm occurred. I discuss this property in detail in the next section.

That said, there is considerable freedom in specifying the P_c^t and consequently, the form of (2.11) is less restrictive than might initially seem. For instance, Fig. 2-3 shows a simple model of a sonar range-finder (like those for which OG mapping was originally developed) that satisfies this form of forward model. Instead of using the reported range directly, the range measurement is decomposed into two measurements from fictitious binary sensors. The first observes cells within a cone shorter than the reported range and registers a non-detection; the second observes a ring of cells centered about the reported range and registers a detection. Part II of this thesis develops several models for chemical plume detection also of the same form. Those models rely on vehicle navigation and measurements of background advective flow in addition to chemical detection to generate the P_c^t .

2.3.1 Inversion and Marginalization

The form of (2.11) is special because it is possible to compute the marginal inverse probabilities $P[m_c | d^t]$ and $P[m_c | \bar{d}^t]$ exactly. Using (2.10) with $z^t \in \{d^t, \bar{d}^t\}$ and

substitutions (2.11) and (2.13) for $P[d^t | \mathbf{m}]$ and $P[\bar{d}^t | \mathbf{m}]$ respectively, leads by iterative evaluation of the sums to

$$P[m_c | d^t] = \frac{1 - (1 - P_F^t)(1 - P_c^t) \prod_{s \neq c}^C (1 - P_s^t P[m_s])}{1 - (1 - P_F^t) \prod_{s=1}^C (1 - P_s^t P[m_s])} P[m_c] \quad (2.14a)$$

$$P[m_c | \bar{d}^t] = \frac{(1 - P_c^t)}{1 - P_c^t P[m_c]} P[m_c] . \quad (2.14b)$$

Note that the large sums over $\{\mathbf{m} : m_c\}$ implied by (2.10) have been replaced with products over C . The derivation of these expressions is provided in App. B.2.

The (information theoretic) entropy of the posterior probability distribution given by (2.14) provides an assessment of the information content of a single measurement from a sensor having the form of (2.11). The entropy of a probability distribution $p(x)$ is defined by (e.g. [118]):

$$H_x = E[-\log_2 p(x)] , \quad (2.15)$$

with conditional entropy defined analogously. The conditional probability distribution $p(\mu_c | \delta^t)$ is bivalued and hence (2.15) resolves to

$$H_{\mu_c | \delta^t}(\delta^t) = -P[m_c | \delta^t] \log_2 P[m_c | \delta^t] - P[\bar{m}_c | \delta^t] \log_2 P[\bar{m}_c | \delta^t] . \quad (2.16)$$

This expression attains a maximum value of 1 for $P[m_c | \delta^t] = 0.5$, corresponding to complete uncertainty. If $P[m_c] = 0.5$, the unconditional entropy of the prior $H_{\mu_c} = 1$ and the posterior conditional entropy can be regarded as an inverse measure of the information content of δ^t . A posterior entropy $H_{\mu_c | \delta^t} = 0$ denotes perfect certainty and hence maximum information. Figure 2-4 shows the posterior entropies $H_{\mu_c | d^t}$ and $H_{\mu_c | \bar{d}^t}$ versus the number of cells observed in a single measurement and versus P_c^t respectively.

Equation (2.14a) links $P[m_c | d^t]$ through the P_s^t to the probabilities that the detection could have been caused by another cell in the map. These alternate possibilities strongly influence the probability that an occupied cell at c was responsible for the detection. In essence, a single detection does not contain sufficient information

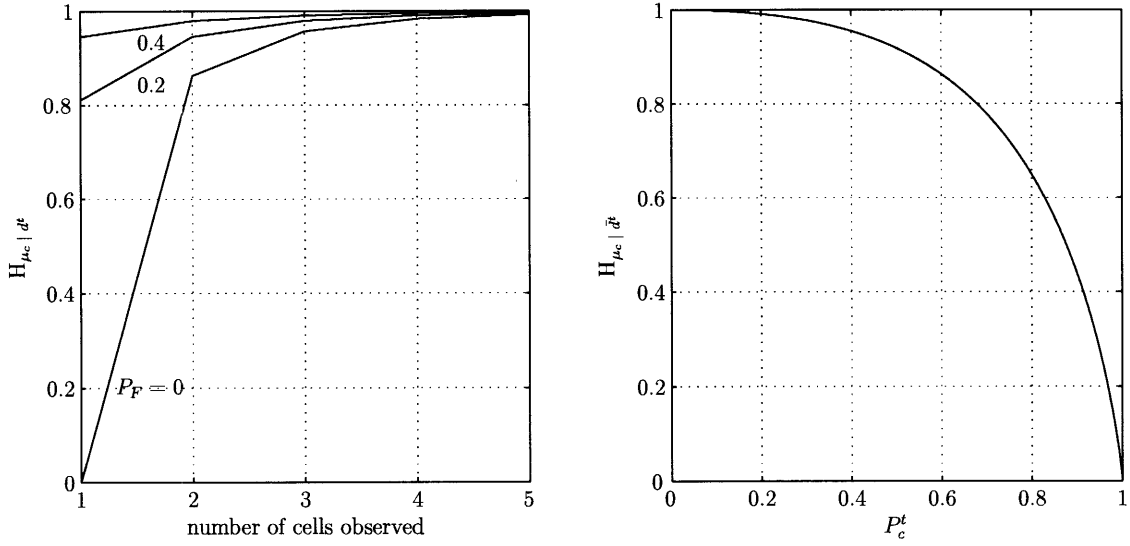


Figure 2-4: Entropy of the inverse sensor model given by (2.14): **(left)** entropy following a detection versus the size of the region observed; **(right)** entropy following a non-detection versus P_c^t . As the number of non-zero P_c^t climbs with the size of the region observed by a detection, the information acquired about any single cell is reduced because a detection is only sufficient to constrain the observed region to containing at least one occupied cell. Information attained from a single non-detection is independent of the number of cells observed. Instead, as probability of detection improves $P_c^t \rightarrow 1$ the posterior entropy monotonically decreases. Fixed values used were $P[m_c] = 0.5$ for both panels, and $P_c^t \equiv 0.5$ for the left panel.

to infer the number of cells that are occupied out of those having non-zero P_c^t , except that at least one must be occupied. Indeed, as the number of non-zero P_s^t climbs with the size of the region observed by the sensor, the information acquired about any single cell is reduced and the inverse probability of occupancy $P[m_c | d^t]$ decays to the prior. This is of course also the case for $P_F^t \rightarrow 1$.

A non-detection, by contrast, does not link $P[m_c | \bar{d}^t]$ to other cells and are easier to interpret. For instance, a non-detection declared by a perfect sensor upon observing some region indicates unequivocally that all cells within the region are unoccupied. The information contained about any single cell in a non-detection is independent of the size of the region observed by the sensor except inasmuch as region size will generally influence P_c^t . Furthermore, as $P_c^t \rightarrow 1$, cell c becomes confidently unoccupied. The probability of false alarm P_F^t plays no role in $P[m_c | \bar{d}^t]$ since clearly no false alarm occurred if the sensor reported a non-detection.

Of course, the whole point of occupancy grid mapping is to combine the informa-

tion available in individual measurements into a consistent map. Detections become valuable when interpreted in light of non-detections that constrain the portions of the map from whence the detection was unlikely to have come.

Notation Some additional notation will facilitate the development in the remainder of this chapter. The set of occupied cells \mathcal{S} in the true map \mathbf{m} has already been defined. Let its complement $\bar{\mathcal{S}}$ denote the set of all vacant cells $\bar{\mathcal{S}} = \{s : \bar{m}_s, s \in \{1, \dots, C\}\}$, and let $S = |\mathcal{S}|$ and $\bar{S} = C - S$ denote the cardinalities of \mathcal{S} and $\bar{\mathcal{S}}$ respectively. Specifying \mathcal{S} or $\bar{\mathcal{S}}$ is equivalent to specifying a map \mathbf{m} . To indicate that the set of all measurements consists of binary detections and non-detections, I use Δ^t instead of Z^t . Further, let D^t denote the set of all measurements that resulted in detections up to time t and the set of all non-detections as \bar{D}^t so that $\Delta^t = \{D^t, \bar{D}^t\}$. Similarly, I denote the set of times that resulted in detections τ^t , and the set of non-detection times $\bar{\tau}^t$. These sets have cardinalities of n^t and $\bar{n}^t = t - n^t$ respectively. Throughout, a normal typeface superscript of t or τ denotes a temporal index. Sets of temporal indices are denoted in boldface and when used as a superscript denote the union over their elements, as in $\Delta^{\tau^t} = \bigcup_{\tau \in \tau^t} \delta^\tau$.

2.3.2 Posterior Independence Property

The fact the P_s^t for all $s \neq c$ do not enter into the inverse model for a single non-detection (2.14b) suggests that independence of the posteriors may persist following any number of non-detections:

$$p(\mathbf{m} \mid \Delta^{\tau^t}) = \prod_{s=1}^C p(\mu_s \mid \Delta^{\tau^t}) . \quad (2.17)$$

If so, then the probability of any particular map \mathbf{m} is simply the product of the marginal posteriors for all occupied and unoccupied cells in that particular map:

$$P[\mathbf{m} \mid \bar{D}^t] = \prod_{s \in \mathcal{S}} P[m_s \mid \bar{D}^t] \cdot \prod_{s \in \bar{\mathcal{S}}} P[\bar{m}_s \mid \bar{D}^t] . \quad (2.18)$$

In fact, it can be shown that a forward model of the form (2.11) is both necessary and sufficient for (2.17) to hold (App. B.3). For $t = 0$, (2.17) merely restates the usual

assumption that the marginal priors are independent. Because of this relationship between $P[m_c | \bar{D}^t]$ and $P[m_c]$, the rest of this thesis adopts the notation $\bar{P}_c^t \triangleq P[m_c | \bar{D}^t]$ with $\bar{P}_c^0 = P[m_c]$.

Equation 2.17 is exploitable inasmuch as it allows recursive computation of the *exact* marginal posteriors when conditioned on exclusively non-detections. In Section 2.4 an exact expression for the odds ratio conditioned on exclusively $\Delta^{\bar{r}^t}$ is found that is identical to the standard OG mapping algorithm. Of course, the information contained in detections must also be incorporated into the OG map for it to be useful.

A natural question is whether the CIM assumption is actually correct when all measurements consist of exclusively non-detections ($\Delta^t = \bar{D}^t$) for a sensor model of the form (2.11). Expressed for exclusively non-detections, the conditional independence assumption required by the standard OG mapping algorithm is $p(\bar{d}^t | \Delta^{\bar{r}^t}, m_c) = p(\bar{d}^t | \mu_c)$. Except under certain trivial conditions, this assumption remains false. A proof exploiting (2.17) is provided in App. B.4; however, intuitive reasoning leads to the same conclusion: Previous non-detections will have made some portions of the map less likely to be occupied than others. Therefore, a measurement including observations of these cells will be less likely to result in a detection than it would have been otherwise, regardless of occupancy of cell c .

2.4 An Exact Algorithm for the Marginal Posteriors

In this section the multiplicative structure of the binary measurement forward model (2.11) is exploited to derive an exact algorithm for the marginal posteriors $P[m_c | D^t, \bar{D}^t]$. Computing the marginal posteriors would nominally require the computing and storing the full posterior $p(\mathbf{m} | D^t, \bar{D}^t)$ followed by marginalization, both daunting tasks considering the large space of all possible maps. The algorithm preserves the dependence between cell posteriors that arises following a detection. It recursively processes non-detections, but requires batch-processing of all detections. Unfortunately, the algorithm is ultimately impracticable as the number of detections n^t increases due to computational and numerical aspects of the batch-processing step. Nevertheless, the

algorithm does enable practical computation of the exact marginals for small n^t , and more significantly, provides the basis for the several approximate algorithms introduced in the next chapter.

In principle the marginal posteriors could be computed for a binary measurement model exactly according to

$$\begin{aligned}
P[m_c \mid D^t, \bar{D}^t] &= \sum_{\mathbf{m}: m_c} p(\mathbf{m} \mid D^t, \bar{D}^t) \\
&= \eta \cdot \sum_{\mathbf{m}: m_c} P[D^t, \bar{D}^t \mid \mathbf{m}] P[\mathbf{m}] \\
&= \eta \cdot \sum_{\mathbf{m}: m_c} \prod_{t \in \tau^t} P[d^t \mid \mathbf{m}] \cdot \prod_{t \in \tau^t} P[\bar{d}^t \mid \mathbf{m}] \cdot P[\mathbf{m}] ,
\end{aligned} \tag{2.19}$$

where η is a normalizing constant and the static world assumption was used to arrive at the last line. This procedure presents analogous implementation problems as inversion of a general forward sensor model, namely storage of $p(\mathbf{m} \mid D^t, \bar{D}^t)$ and costly 2^C -term summation over the set of all maps with cell c occupied $\{\mathbf{m} : m_c\}$.

As with the inversion of $P[d^t \mid \mathbf{m}]$ and $P[\bar{d}^t \mid \mathbf{m}]$ in the previous section, the special form of (2.11) enables such marginalization by direct summation to be circumvented. This leads to a particularly simple and computationally attractive (recursive) expression for the marginal posterior odds conditioned on non-detections exclusively $\bar{\rho}_c^t$:

$$\begin{aligned}
\bar{\rho}_c^t &\triangleq P[m_c \mid \bar{D}^t] / P[\bar{m}_c \mid \bar{D}^t] \\
&= (1 - P_c^t) \bar{\rho}_c^{t-1}
\end{aligned} \tag{2.20}$$

The simplicity of this result is perhaps not surprising given property (2.17). Interestingly it is also identical to the odds ratio computed via the standard algorithm for $\Delta^t = \bar{D}^t$; however, the CIM is not required to derive it for forward sensor models of the form (2.11). The proof appears in App. B.5.

The expression for the general odds ratio $r_c^t \triangleq P[m_c \mid D^t, \bar{D}^t] / P[\bar{m}_c \mid D^t, \bar{D}^t]$ does not simplify as cleanly; however, it is still possible to avoid direct marginalization and its attendant exponential cost in map size in favor of a non-recursive expression whose computational cost instead scales linearly in map size but exponentially with

the number of detections n^t :

$$\rho_c^t = \frac{1 + \sum_{q=1}^{n^t} (-1)^q \sum_{\boldsymbol{\eta} \subseteq \boldsymbol{\tau}^t, \boldsymbol{\eta} \in \mathbb{N}^q} \prod_{i=1}^q (1 - P_c^{\eta_i}) f_c^t(\boldsymbol{\eta})}{1 + \sum_{q=1}^{n^t} (-1)^q \sum_{\boldsymbol{\eta} \subseteq \boldsymbol{\tau}^t, \boldsymbol{\eta} \in \mathbb{N}^q} f_c^t(\boldsymbol{\eta})} \quad (2.21)$$

$$r_c^t = \rho_c^t \cdot \bar{\rho}_c^t.$$

In the above $f_c^t(\boldsymbol{\eta})$ denotes the product

$$f_c^t(\boldsymbol{\eta}) = \prod_{i=1}^{|\boldsymbol{\eta}|} (1 - P_F^{\eta_i}) \prod_{s \neq c} \left(1 - \left(1 - \prod_{i=1}^{|\boldsymbol{\eta}|} (1 - P_s^{\eta_i}) \right) \bar{P}_s^t \right). \quad (2.22)$$

The limit $\boldsymbol{\eta} \subseteq \boldsymbol{\tau}^t, \boldsymbol{\eta} \in \mathbb{N}^q$ denotes all q -length combinations of the detection times $\tau_i \in \boldsymbol{\tau}^t$. The upper limits on the products in (2.22) refer to the cardinality of $\boldsymbol{\eta}$ (i.e. q in (2.21)). At $t = 0$, r_c^0 and $\bar{\rho}_c^0$ are initialized as $r_c^0 = \bar{\rho}_c^0 = \frac{P[m_c]}{1 - P[m_c]}$.⁶ Eq. (2.21) is complicated because it maintains the dependencies between the marginal posteriors that form after cells have been observed by more than one measurement that resulted in a detection. The derivations of (2.20) and (2.21) are provided in App. B.5.

Because of the CIM assumption required to derive it, the standard OG mapping algorithm fails to maintain the dependencies that arise between cells following detections, and consequently it fails to produce correct values for the marginal posteriors. Eq. (2.21) enables exact computation of the posteriors for maps of arbitrary size, albeit for only a few detections. Figure 2-5 demonstrates the improvement attained by the exact algorithm over the standard algorithm for a detection followed by a non-detection on a simple 2-cell map. The next chapter explores the consequences of the standard conditional independence assumption on realistically sized maps. Before considering the practical application of (2.21), the next section quantifies the computational cost incurred by computing the posteriors via (2.21) and the numerical stability of the result. Unfortunately, (2.21) is not a panacea; however, it provides the foundation for the several novel approximate algorithms introduced in the next

⁶Note that $\rho_c^t \neq P[m_c | D^t] / P[\bar{m}_c | D^t]$. Rather, the definition in (2.21) reflects the parallel roles of ρ_c^t and $\bar{\rho}_c^t$ in the algorithmic implementations of (2.21) presented in this chapter and the next. A useful definition for ρ_c^t is $\rho_c^t \triangleq P[D^t | \bar{D}^t, m_c] / P[D^t | \bar{D}^t, \bar{m}_c]$.

Algorithm 1 An exact algorithm for the odds ratio. The required initialization is $\bar{\rho}_c^0 = \bar{P}_c^t / (1 - \bar{P}_c^t)$. Unless otherwise noted, all subscripts c imply the operation is for all $c \in \{1, \dots, C\}$.

Require:

$$\begin{aligned} &\bar{\rho}_c^{t-1} \\ &P_c^t \\ &P_c^{\tau_i}, \forall \tau_i \in \tau^{t-1} \end{aligned}$$

```

 $r_c^t = \text{og\_exact}\{$ 
  1: if  $d^t$  (detection) then
  2:    $\bar{\rho}_c^t = \bar{\rho}_c^{t-1}$ 
  3:    $\bar{P}_c^t = \bar{\rho}_c^t / (1 + \bar{\rho}_c^t)$ 
  4: else if  $\bar{d}^t$  (nondetection) then
  5:    $\bar{\rho}_c^t = (1 - P_c^t) \bar{\rho}_c^{t-1}$ 
  6:    $\bar{P}_c^t = \bar{\rho}_c^t / (1 + \bar{\rho}_c^t)$ 
  7: end if
  8:  $\rho_c^t = \frac{1 + \sum_{q=1}^{n^t} (-1)^q \sum_{\eta \subseteq \tau^t, \eta \in \mathbb{N}^q} \prod_{i=1}^q (1 - P_c^{\eta_i}) f_{\mathcal{L}}^q(\eta)}{1 + \sum_{q=1}^{n^t} (-1)^q \sum_{\eta \subseteq \tau^t, \eta \in \mathbb{N}^q} f_{\mathcal{L}}^q(\eta)}$ 
  9:  $r_c^t = \rho_c^t \cdot \bar{\rho}_c^t$ 
 $\}$ 

```

chapter.

Algorithm 1 implements the iterative update of an OG map via equations (2.20) and (2.21) in pseudocode. An implementation exploiting the lack of dependence between detections that observe portions of the map not related to one another either by direct overlap or through overlap with shared neighbors is provided in App. C. The implementation can cut down significantly on the computational load if detections tend to occur in disconnected clusters since such groups of detections can be treated independent of one another without approximation.

2.4.1 Numerical and Computational Considerations

The number of arithmetic operations required to compute ρ_c^t via (2.21) is approximately proportional⁷ to $Cn^t2^{n^t}$. That is, the computational cost is linear in map size

⁷As written, the precise number of arithmetic operations required by (2.21) is $6(C+1) + 4n^t + 6(C-1)2^{n^t} + (2C+1)n^t2^{n^t}$. An upper bound for this number is $14Cn^t2^{n^t}$, thus the proportionality

	Std. Algorithm	Exact Alg.
$t = 0$: prior	<div style="display: inline-block; border: 1px solid black; padding: 5px; margin-right: 10px;"> ^A \bar{P} </div> <div style="display: inline-block; border: 1px solid black; padding: 5px;"> ^B \bar{P} </div>	<div style="display: inline-block; border: 1px solid black; padding: 5px; margin-right: 10px;"> ^A \bar{P} </div> <div style="display: inline-block; border: 1px solid black; padding: 5px;"> ^B \bar{P} </div>
$t = 1$: detection observes cells A & B $D^1 = \{d^1\}, \bar{D}^1 = \emptyset$	<div style="display: inline-block; border: 2px solid red; padding: 5px; margin-right: 10px;"> ^A $\frac{1}{2-P}$ </div> <div style="display: inline-block; border: 2px solid red; padding: 5px;"> ^B $\frac{1}{2-P}$ </div>	<div style="display: inline-block; border: 2px solid red; padding: 5px; margin-right: 10px;"> ^A $\frac{1}{2-P}$ </div> <div style="display: inline-block; border: 2px solid red; padding: 5px;"> ^B $\frac{1}{2-P}$ </div>
$t = 2$: nondetection observes cell A only $D^2 = \{d^1\}, \bar{D}^2 = \{d^2\}$	<div style="display: inline-block; border: 2px solid blue; padding: 5px; margin-right: 10px;"> ^A 0 </div> <div style="display: inline-block; border: 1px solid black; padding: 5px;"> ^B $\frac{1}{2-P}$ </div>	<div style="display: inline-block; border: 2px solid blue; padding: 5px; margin-right: 10px;"> ^A 0 </div> <div style="display: inline-block; border: 1px solid black; padding: 5px;"> ^B 1 </div>

Figure 2-5: Comparison of standard OG mapping algorithm with the exact algorithm specified by (2.21) and (2.20) for a two-cell map. A perfect receiver is assumed: $P_c^t \equiv 1$, $P_F^t \equiv 0$. The values within each cell indicate the posterior probability of occupancy after processing the data from each time step. Because a perfect receiver cannot miss a detection, both algorithms correctly assign a zero probability of occupancy to cell A after measuring the state of cell A (exclusively) at $t = 2$. At the same time the exact algorithm also correctly identifies cell 2 as conclusively occupied, whereas the standard algorithm does not update its estimate for cell B.

C and exponential in the number of detections n^t . Figure 2-6 shows the results of a numerical confirmation of this scaling. Direct application of (2.21) to update all cells would incur quadratic cost in C ; however, the $f_c^t(\boldsymbol{\eta})$ need only be computed once per iteration so that linear cost in C is maintained. While exponential cost in anything is undesirable, it is possible to use (2.21) to compute the exact odds ratios for small n^t . This is not the case in general because of the exponential cost (2^C) in map size implied by the marginalization in (2.19).

Computation of the exact odds via (2.21) can be numerically problematic as well, particularly if the P_c^t and \bar{P}_c^t are small. Under these conditions, the products within the inner sums in the numerator and denominator of (2.21) tend to equal values near 1. Summation over all q -length combinations of $\tau_i \in \boldsymbol{\tau}^t$ therefore tends to yield values near $\binom{n^t}{q}$ for the inner sums. The outer summation over the resulting alternating series then tends to yield values close to zero.⁸ Numerical investigation suggests rapid convergence of the sums in the numerator and denominator of (2.21) toward zero as n^t is increased. Indeed, the numerator and denominator of ρ_c^t are $P[D^t | \bar{D}^t, m_c]$ and

claimed in the main text is conservative.

⁸By the Binomial Theorem, $\sum_{q=0}^{n^t} (-1)^q \binom{n^t}{q} = 0$.

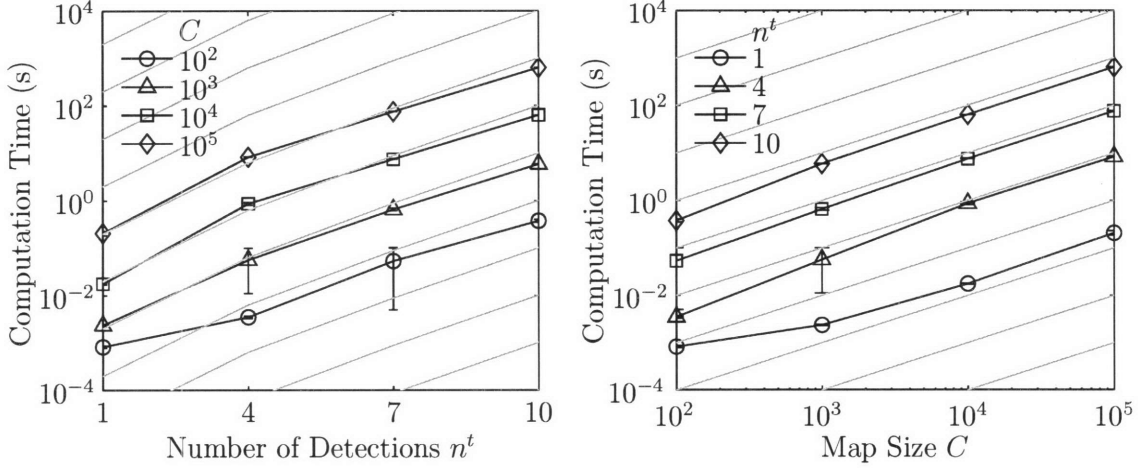


Figure 2-6: Computation time for an implementation of Algorithm 1 versus the number of detections n^t (left) and the size of the map C (right). Each data point represents the average of 10 runs. The gray lines in both plots are curves of $y = \text{const} \cdot C n^t 2^{n^t}$ for a range leading constants. These plots confirm a computational complexity that scales approximately as $n^t 2^{n^t}$ with n^t and linearly with C . Deviation from this scaling for small maps and low n^t is due to overhead.

$P[D^t \mid \bar{D}^t, \bar{m}_c]$ respectively. The probability that the set of measurement times τ^t resulted in detections will approach zero as $t \rightarrow \infty$ in the same manner that the probability of observing any particular sequence of heads and tails from a sequence of coin flips will approach zero as the number of flips increases; however, the convergence is particularly sensitive to the number of detections in this case because the probability of observing sequences containing detections is unlikely for small P_c^t and \bar{P}_c^t . This behavior becomes problematic when the results of the sums approach the machine precision of the much larger $\geq O(1)$ terms within them.

The much simplified case of a 2-cell map with repeated identical and uniform measurements ($C = 2$; $P_c^t \equiv P_D$; $\bar{P}_c^t \equiv P_0$; $P_F^t \equiv 0$) illustrates the behavior. Under these conditions and for a sequence of n^t detections, numerical failure will occur first in the denominator of (2.21), which reduces to

$$\sum_{q=0}^{n^t} (-1)^q \binom{n^t}{q} (1 - (1 - (1 - P_D)^q) P_0) . \quad (2.23)$$

Upon simplification and application of the Binomial Theorem, the above becomes simply $(P_D)^{n^t} P_0$ which converges exponentially toward zero, and very rapidly for small

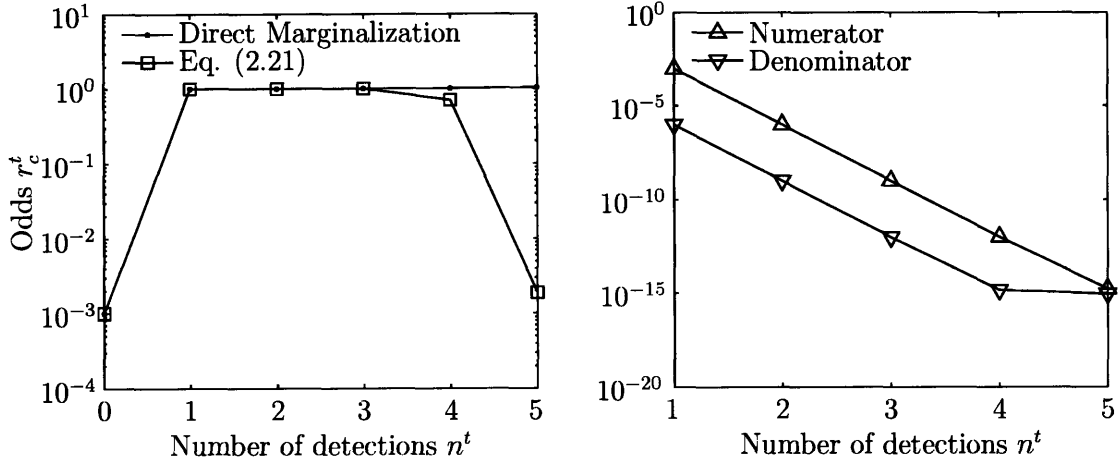


Figure 2-7: Comparison of the odds ratio computed from (2.21) versus via direct marginalization for a two-cell map (left) and the values of the numerator and denominator of (2.21) as functions of the number of detections. The precision of $O(1)$ numbers for these computations was approximately 10^{-16} . Numerical failure of (2.21) is evident at $n^t = 4$, consistent with finite precision problems in representing the denominator of (2.21) for $P_c^t \equiv 0.001$; $\bar{P}_c^t \equiv 0.001$; $P_F^t \equiv 0$. These values were chosen for illustrative purposes and not as representative of a real application.

P_D . Numerical failure will occur when the product $(P_D)^{n^t} P_0$ approaches the machine precision of the $O(1)$ summands in (2.23). Figure 2-7 illustrates this by comparison to the solution computed from marginalization over the space of all maps:

$$r_c^t = \frac{(P_D)^{n^t} (1 - P_0) P_0 + (1 - (1 - P_D)^2)^{n^t} P_0^2}{(P_D)^{n^t} (1 - P_0) P_0}. \quad (2.24)$$

Direct marginalization is possible here because of the small size of the map.

In situations where P_F^t is not negligible, the numerical results produced by (2.21) remain accurate for larger n^t . This is because larger P_F^t have the effect of dampening the oscillations of the outer alternating sums in the numerator and denominator. Non-zero P_F^t also relax the constraint that data be completely consistent, which lends some intuitive support to the improved numerical behavior of (2.21) for larger P_F^t .

Finally, it should be noted that the standard Bayesian OG algorithm as usually implemented works by summing the logarithms of the individual factors of (2.5) so as to avoid numerical problems associated with representing values near 1 and 0 [85,118]. Because (2.21) consists of both sums and products, logarithms are of limited utility

here.

2.5 Summary, Contributions, and Future Work

This chapter introduced the standard Bayesian OG mapping algorithm, the key assumptions required to derive it, and paid particular attention to the assumed conditional independence of measurements (CIM assumption) and its manifestation in OG maps. Recall that the CIM states that measurements are conditionally independent of one another given knowledge of the state of a single cell in the map. The assumption applies to each cell updated by the algorithm at each iteration. Though the CIM assumption enables the mapping problem to be decomposed into a collection of one-dimensional binary state-estimation problems, the results do not reflect the dependencies between cell states that naturally arise as a consequence of using sensors that sweep over multiple cells [117]. I argued that such dependencies are particularly important in environments with a naturally low density of occupied cells, where explanations of the data that require a minimal number of occupied cells should be strongly favored.

To address this shortcoming of the standard method, an exact expression for the marginal posteriors was derived that is applicable to a special form of invertible binary forward sensor model. It was shown that non-detections could be incorporated into the OG map recursively, with computational cost equivalent to the standard algorithm. At present, incorporating detections requires batch-processing of all detections, with computational cost exponential in the number of detections. Nevertheless, the existence of this exact expression for the marginals proves that (for a specific class of sensor model) the exponential complexity in map size suggested by direct marginalization of the full posterior $p(\mathbf{m} \mid D^t, \bar{D}^t)$ can be exchanged for exponential complexity in the number of detections. For small numbers of detections and arbitrary numbers of non-detections, this work enables the practical attainment of exact OG maps regardless of map size.

The results presented herein differ from previous research aimed at compensating for the consequences of the CIM assumption in OG maps of indoor environments. To my knowledge, the only other directly comparable exact method was developed by Berler and Shimony [10]. Though they employ an OG only for display purposes

(instead regions are dynamically defined with each new measurement), their algorithm is capable of producing exact posteriors, and like the exact methods presented in this chapter (§ 2.4), theirs scales exponentially in computational complexity with the number of measurements. Unlike in their algorithm, the complexity of the environmental representation (the map) remains fixed at C grid cells for the algorithms developed in this chapter and the next. In contrast to methods employing augmented OGs (principally to address specular reflections in indoor environments, e.g. [48,58]), the CIM assumption is addressed without augmenting the representation of the environment. Finally, the present approach has been formulated utilizing a sensor model of a particular form, but not specific to a particular sensing modality (e.g. the “context sensitive” OG mapping algorithm developed for sonar range finders by Moravec and Cho [86]).

Chapter 3

Approximate Algorithms for Low Prior Environments

Simulations presented in this chapter demonstrate the significant degradation in accuracy of the single cell posteriors that occurs when the standard OG mapping algorithm is applied to an environment with a low prior probability of occupancy. The last chapter provided an exact expression (2.21) for the marginal posterior odds, but unfortunately also revealed numerical and computational aspects of that expression that rendered it ultimately impractical. This chapter presents several novel approximate algorithms that each retain some ability to maintain the dependence that arises between cell states after observations that result in detections.

Two of these new algorithms are closely related to the standard algorithm. This first class requires variants of the standard CIM assumption that leverage the utility of (2.21) for small numbers of detections. As was shown in the preceding chapter, the marginal posteriors conditioned on exclusively non-detections maintain their independence (§ 2.3.2). Since non-detections are easy to handle exactly, I focus on developing relaxed assumptions on the independence of times that resulted in detections. A second class of algorithms instead requires assuming independence of the marginal posteriors. Algorithms of this class essentially “revise the prior” to reflect the current posterior before incorporating a new measurement. These algorithms are also amenable to extension via the processing of subsets of detections with (2.21).

In each case, analytic sufficient conditions for the algorithm to produce exact results are provided. The performance of each algorithm is then evaluated in a sim-

ulated low prior environment. The simulation results show qualified improvement over the standard algorithm, with the best results produced by algorithms assuming independence of the marginal posteriors. The principal improvement demonstrated is the computation of posteriors more consistent with the prior, particularly a reduced propensity to produce very high, unrealistic posteriors over large portions of the map.

3.1 Approximate Algorithms from Relaxed CIM Assumptions

The various extensions of the CIM assumption are perhaps best understood by rewriting the odds ratio $r_c^t \triangleq \text{P} [m_c \mid D^t, \bar{D}^t] / \text{P} [\bar{m}_c \mid D^t, \bar{D}^t]$ in a modified form with the odds ratio conditioned on exclusively non-detections $\bar{\rho}_c^t$ as a factor:

$$\begin{aligned} r_c^t &= \frac{\text{P} [D^t \mid \bar{D}^t, m_c]}{\text{P} [D^t \mid \bar{D}^t, \bar{m}_c]} \cdot \frac{\text{P} [m_c \mid \bar{D}^t]}{\text{P} [\bar{m}_c \mid \bar{D}^t]} \\ &= \frac{\text{P} [D^t \mid \bar{D}^t, m_c]}{\text{P} [D^t \mid \bar{D}^t, \bar{m}_c]} \cdot \bar{\rho}_c^t. \end{aligned} \tag{3.1}$$

Eq. (3.1) is readily derived by the application of Bayes Rule to the definition of r_c^t . The key quantity in (3.1) is the ratio of the probabilities of the measurements that resulted in detections given a single cell c and all non-detections $\text{P} [D^t \mid \bar{D}^t, m_c] / \text{P} [D^t \mid \bar{D}^t, \bar{m}_c]$. Computationally tractable expressions for this quantity can be attained by placing assumptions on the conditional independence of measurements at times $\tau \in \boldsymbol{\tau}^t$

$$\text{p} \left(\Delta^{\boldsymbol{\tau}^t} \mid \Delta^{\bar{\boldsymbol{\tau}^t}}, \mu_c \right).$$

These are the measurements that resulted in detections. Obviously such assumptions cannot be made a priori; however, the output of approximate algorithms can be qualitatively assessed by analyzing the quality of the assumptions required given the measurement history. There are no guarantees, but bad assumptions usually lead to bad maps.

For instance, the standard OG mapping algorithm computes the odds ratio conditioned on non-detections exclusively, $\bar{\rho}_c^t$, exactly. This fact could not be deduced from

the derivation alone if the standard CIM assumption for binary measurements were left in the unaltered form (2.2) where it applies to both detections and non-detections. The next section presents a modified, weaker version of the CIM assumption that nonetheless leads to the standard algorithm, thereby confirming directly that the $\bar{\rho}_c^t$ computed via the standard algorithm are exact.

The standard CIM assumption for binary measurements

The standard CIM assumption expressed for binary measurements is

$$p(\delta^t \mid \Delta^{t-1}, \mu_c) = p(\delta^t \mid \mu_c) , \quad (3.2)$$

where equality must hold for all t . By repeated application of the definition of conditional independence, (3.2) is equivalent to

$$p(\Delta^t \mid \mu_c) = \prod_{\tau=1}^t p(\delta^\tau \mid \mu_c) , \quad (3.3)$$

which states that all measurements carry no information about one another not subsumed by knowledge of the state of the single cell c . This assumption does not apply globally across all cells; that is, only the version corresponding to cell c is required each time cell c is updated.

With a forward sensor model of the form (2.11), however, the standard algorithm can be derived by making the weaker (necessary, but not sufficient) assumption

$$p(\Delta^{\tau^t} \mid \Delta^{\bar{\tau}^t}, \mu_c) = \prod_{\tau \in \tau^t} p(\delta^\tau \mid \mu_c) \quad (3.4)$$

which removes measurement times that resulted in non-detections from (3.3), though it still requires that all measurements at times that resulted in detections be conditionally independent of measurements at times that resulted in non-detections. The derivation is carried out below in § 3.1.1.

Relaxing CIM by Grouping Measurements

Intuition suggests that results superior to the standard OG mapping algorithm might be achievable by requiring that conditional independence hold only between subsets of detections, rather than between each detection, that is, by further relaxing (3.4). Such an approach is made possible by exploiting (2.21) to compute the exact odds ratio for subsets of correlated detections. These subsets must be sufficiently small to avoid the numerical and computational pitfalls of (2.21) discussed in § 2.4.1.

Managing subsets of detection times requires introducing some notation. Let $\mathbf{g}_k^t \subseteq \boldsymbol{\tau}^t$ denote the k^{th} subset of detection times from a partition of $\boldsymbol{\tau}^t$. Also, let \mathcal{G}^t denote the set of indices k such that

$$\bigcup_{k=1}^{|\mathcal{G}^t|} \mathbf{g}_k^t = \boldsymbol{\tau}^t \quad (3.5)$$

where $|\mathcal{G}^t|$ is the cardinality of \mathcal{G}^t or equivalently its maximum element. It will prove convenient to also define a subset of indices \mathcal{G}_F^t into groups of some maximum size:

$$\mathcal{G}_F^t \triangleq \{k: k \in \mathcal{G}^t, |\mathbf{g}_k^t| = \text{MAX_GROUP_MEMBERS}\} . \quad (3.6)$$

With these definitions, two weaker variations on (3.4), again in the sense of being necessary but not sufficient, are

$$p\left(\Delta^{\boldsymbol{\tau}^t} \mid \Delta^{\bar{\boldsymbol{\tau}}^t}, \mu_c\right) = \prod_{k=1}^{|\mathcal{G}^t|} p\left(\Delta^{\mathbf{g}_k^t} \mid \mu_c\right) \quad (3.7a)$$

$$p\left(\Delta^{\boldsymbol{\tau}^t} \mid \Delta^{\bar{\boldsymbol{\tau}}^t}, \mu_c\right) = \prod_{k=1}^{|\mathcal{G}^t|} p\left(\Delta^{\mathbf{g}_k^t} \mid \Delta^{\bar{\boldsymbol{\tau}}^t}, \mu_c\right) \quad (3.7b)$$

The first of these collapses to (3.4) if $|\mathbf{g}_k^t| = 1 \ \forall \ k$, and like the standard assumption, assumes that measurements at times that resulted in non-detections carry no information relevant to the outcome of the remaining measurements not subsumed by knowledge of μ_c . For this reason, I regard the algorithm derived from (3.7a) in (§ 3.1.2) as an extension of the standard algorithm. The second assumption above further relaxes (3.4) by retaining the conditioning on $\Delta^{\bar{\boldsymbol{\tau}}^t}$. I call the novel algorithm

that results from this assumption the Conditional Independence of Detections (CID) algorithm. Its derivation is given in § 3.1.3.

As with the standard CIM assumption, assumptions (3.7a & b) are unique to each cell c . Essentially, the algorithms with these assumptions at their core still treat the problem of estimating the marginal posteriors as a of set C independent binary estimation problems; albeit with some all measurements properly regarded correlated observations observations.

Furthermore, these modified CIM assumptions beg the question of how $\boldsymbol{\tau}^t$ should be partitioned into the subsets \boldsymbol{g}_k^t . Numerical examples presented later in this chapter (§ 3.3) indicate the numerical value of the posterior varies substantially for different partitions of $\boldsymbol{\tau}^t$. In fact, since the assumption required is unique to each cell, a different partition could be used for each cell; however, the computational burden of updating the whole map is greatly decreased by using a global partition.

The results presented in this work rely on a heuristic method for partitioning $\boldsymbol{\tau}^t$. The method, implemented in pseudocode as Algorithm 2, uses pairwise comparisons of detections to assign new detections to existing subsets or else to instantiate a new subset. The algorithm groups relatively correlated measurements into the same subset so as to avoid treating them as independent. The normalized correlation coefficient $v_{t,\tau}$ between a new measurement δ^t and an old measurement δ^τ is

$$\begin{aligned} v_{t,\tau} &\triangleq \frac{\text{cov}(\delta^t, \delta^\tau)}{\sqrt{\text{var}(\delta^t) \text{var}(\tau)}} \\ &= \frac{f^0(t, \tau) - f^0(t) f^0(\tau)}{\left((1 - f^0(t)) f^0(t) (1 - f^0(\tau)) f^0(\tau) \right)^{\frac{1}{2}}}, \end{aligned} \quad (3.8)$$

where the $f^0(\boldsymbol{\eta})$ for $\boldsymbol{\eta} \subseteq \{t, \tau\}$ are given by

$$f^0(\boldsymbol{\eta}) = \prod_{i=1}^{|\boldsymbol{\eta}|} (1 - P_F^{\eta_i}) \prod_{s=1}^C \left(1 - \left(1 - \prod_{i=1}^{|\boldsymbol{\eta}|} (1 - P_s^{\eta_i}) \right) \bar{P}_s^0 \right). \quad (3.9)$$

The correlation coefficient between two detections provides a measure of how dependent they are. Overlapping measurements will tend to be highly correlated, whereas measurements that observe different portions of the map are independent and therefore uncorrelated. Highly correlated measurements strongly violate the CIM assump-

tion and should be processed exactly if possible.

Eq. (3.9) differs from (2.22) only in the limits of the product over cell indices. Note that $v_{t,\tau} \geq 0$. If the largest $v_{t,\tau}$ so computed surpasses a fixed threshold, then δ^t is added to the subset of τ^t to which δ^r belongs, or else a new subset is instantiated. Once a subset has reached the maximum size set by MAX_GROUP_MEMBERS, it becomes static. New measurements that might have been grouped with members of a static set are either placed into other subsets or else instantiate new subsets.

At the beginning of this section it was suggested that mapping algorithms that make use of (2.21) to compute the exact odds ratio for subsets of measurements should produce progressively better approximations to the exact posteriors as subset size increases. Though intuitively attractive, there is no guarantee that just weakening the standard assumption according to the relaxed CIM assumptions (3.7a) or (3.7b) will result in an improved map. Indeed, simulation results in § 3.3 show a more complex picture. Before presenting these, I derive the binary OG mapping algorithms that result from making the CIM-like assumptions introduced above. The starting point in each case is (3.1). The derivations are followed by discussions of the computational features of each algorithm. Pseudocode implementations of each algorithm are also provided.

3.1.1 Standard Algorithm for Binary Measurements

To derive the standard algorithm from (3.1), apply (3.4) and then Bayes Rule:

$$\begin{aligned}
r_c^t &= \frac{P[D^t \mid \bar{D}^t, m_c]}{P[D^t \mid \bar{D}^t, \bar{m}_c]} \bar{\rho}_c^t \\
&= \frac{\prod_{\tau \in \tau^t} P[d^\tau \mid m_c]}{\prod_{\tau \in \tau^t} P[d^\tau \mid \bar{m}_c]} \cdot \bar{\rho}_c^t \\
&= \prod_{\tau \in \tau^t} \frac{P[m_c \mid d^\tau] P[\bar{m}_c]}{P[\bar{m}_c \mid d^\tau] P[m_c]} \cdot \bar{\rho}_c^t \\
&= \left(\frac{1 - P[m_c]}{P[m_c]} \right)^{n^t} \cdot \prod_{\tau \in \tau^t} \frac{P[m_c \mid d^\tau]}{1 - P[m_c \mid d^\tau]} \cdot \bar{\rho}_c^t.
\end{aligned} \tag{3.10}$$

Algorithm 2 Pairwise grouping algorithm.

Require:

P_c^t
 $P_c^{\tau_i}, \forall \tau_i \in \tau^{t-1}$
 $\mathcal{G}^{t-1}, \mathcal{G}_F^{t-1}$
 $\mathbf{g}_k^{t-1}, \forall k \in \mathcal{G}^{t-1}$
1: Loop over existing subsets:
2: **for all** $\{k: k \in \mathcal{G}^{t-1}, k \notin \mathcal{G}_F^{t-1}\}$ **do**
3: Loop over members of each subset:
4: **for all** $\{j: \tau_j \in \mathbf{g}_k^{t-1}\}$ **do**
5: $w_j = \frac{f^{Q(t, \tau_j)} - f^{Q(t)} f^{Q(\tau_j)}}{\left((1 - f^{Q(t)}) f^{Q(t)} (1 - f^{Q(\tau_j)}) f^{Q(\tau_j)} \right)^{\frac{1}{2}}}$
6: **end for**
7: $v_k = \max_j w_j$
8: **end for**
9: $k_{max} = \operatorname{argmax}_k v_k$
10: **if** $v_{k_{max}} \geq \text{GROUPING_THRESHOLD}$ **then**
11: Add to existing subset:
12: $\mathbf{g}_k^t = \mathbf{g}_k^{t-1} \forall k \in \mathcal{G}^{t-1}, k \neq k_{max}$
13: $\mathcal{G}^t = \mathcal{G}^{t-1}$
14: $\mathbf{g}_{k_{max}}^t = \{\mathbf{g}_{k_{max}}^{t-1}, t\}$
15: **if** $|\mathbf{g}_{k_{max}}^t| = \text{MAX_GROUP_MEMBERS}$ **then**
16: Make subset static:
17: $\mathcal{G}_F^t = \{\mathcal{G}_F^{t-1}, k_{max}\}$
18: **else**
19: $\mathcal{G}_F^t = \mathcal{G}_F^{t-1}$
20: **end if**
21: **else**
22: Instantiate new group:
23: $\mathbf{g}_k^t = \mathbf{g}_k^{t-1} \forall k \in \mathcal{G}^{t-1}$
24: $\mathcal{G}^t = \{\mathcal{G}^{t-1}, |\mathcal{G}^{t-1}| + 1\}$
25: $\mathbf{g}_{|\mathcal{G}^t|}^t = \{t\}$
26: **end if**

To see that (3.10) is equivalent to the recursive form of the standard algorithm (2.5) for $z^t = \delta^t$ requires some manipulation. From (2.20),

$$\begin{aligned}
\bar{\rho}_c^t &= \prod_{\tau \in \bar{\tau}^t} (1 - P_c^\tau) \cdot \frac{\text{P}[m_c]}{1 - \text{P}[m_c]} \\
&= \left(\frac{1 - \text{P}[m_c]}{\text{P}[m_c]} \right)^{\bar{n}^t} 75 \prod_{\tau \in \bar{\tau}^t} \frac{\text{P}[m_c \mid \bar{d}^\tau]}{1 - \text{P}[m_c \mid \bar{d}^\tau]}, \tag{3.11}
\end{aligned}$$

Algorithm 3 The standard algorithm for approximating the odds ratio. The required initializations are $\bar{\rho}_c^0 = \frac{\bar{P}_c^0}{1-\bar{P}_c^0}$ and $\rho_c^0 = 1$. This implementation is somewhat atypical to facilitate comparison with the other algorithms introduced in this work. In particular, separate maintenance of ρ_c^t and $\bar{\rho}_c^t$ is not necessary.

Require:

```

     $t > 0$ 
     $\rho_c^{t-1}, \bar{\rho}_c^{t-1}$ 
     $P_c^t, P_F^t$ 
1: if  $d^t$  (detection) then
2:    $\bar{\rho}_c^t = \bar{\rho}_c^{t-1}$ 
3:    $P[m_c \mid d^t] = \frac{1-(1-P_F^t)(1-P_c^t) \prod_{s \neq c}^C (1-P_s^t P[m_s])}{1-(1-P_F^t) \prod_{s=1}^C (1-P_s^t P[m_s])} P[m_c]$ 
4:    $\rho_c^t = \frac{P[m_c \mid d^t]}{1-P[m_c \mid d^t]} \cdot \rho_c^{t-1}$ 
5: else if  $\bar{d}^t$  (nondetection) then
6:    $\bar{\rho}_c^t = (1-P_c^t) \bar{\rho}_c^{t-1}$ 
7:    $\rho_c^t = \rho_c^{t-1}$ 
8: end if
9:  $r_c^t = \frac{1-\bar{P}_c^0}{\bar{P}_c^0} \rho_c^t \cdot \bar{\rho}_c^t$ 

```

where the first line can be attained from the second by substituting (2.14b) for $P[m_c \mid \bar{d}^t]$ and simplifying. Combining these results yields

$$\begin{aligned}
 r_c^t &= \left(\frac{1-P[m_c]}{P[m_c]} \right)^t \cdot \prod_{\tau=1}^t \frac{P[m_c \mid \delta^\tau]}{1-P[m_c \mid \delta^\tau]} \cdot \frac{P[m_c]}{1-P[m_c]} \\
 &= \frac{1-P[m_c]}{P[m_c]} \cdot \frac{P[m_c \mid \delta^t]}{1-P[m_c \mid \delta^t]} \cdot r_c^{t-1}.
 \end{aligned} \tag{3.12}$$

For completeness, Algorithm 3 implements an equivalent recursive form in pseudocode.

This algorithm “does the right thing” following a detection, in the sense of increasing the posterior in cells where $P[m_c \mid d^t]$ is high; however it does so without respect to information in previous measurements about the rest of the map. Simulations presented later show this tends to lead to inflated estimates of the posteriors over large areas. The advantage of this algorithm is its ease of implementation and

minimal memory and processing requirements.

3.1.2 Extended-Standard Algorithm

The relaxed CIM assumption (3.7a) requires conditional independence to hold only between subsets of detections and leads to an extended version of the standard algorithm. Starting from (3.1), apply (3.7a), followed by Bayes Rule:

$$\begin{aligned}
r_c^t &= \frac{\text{P} [D^t \mid \bar{D}^t, m_c]}{\text{P} [D^t \mid \bar{D}^t, \bar{m}_c]} \cdot \bar{\rho}_c^t \\
&= \prod_{k=1}^{|\mathcal{G}^t|} \frac{\text{P} [D^{\mathbf{g}_k^t} \mid m_c]}{\text{P} [D^{\mathbf{g}_k^t} \mid \bar{m}_c]} \cdot \bar{\rho}_c^t \\
&= \prod_{k=1}^{|\mathcal{G}^t|} \frac{\text{P} [m_c \mid D^{\mathbf{g}_k^t}] (1 - \bar{P}_c^0)}{\text{P} [\bar{m}_c \mid D^{\mathbf{g}_k^t}] \bar{P}_c^0} \cdot \bar{\rho}_c^t .
\end{aligned} \tag{3.13}$$

To evaluate the arguments of the product above, use the expressions for $\text{P} [m_c \mid D^t, \bar{D}^t]$ and $\text{P} [\bar{m}_c \mid D^t, \bar{D}^t]$ given by equations (B.29) and (B.30) with $\bar{D}^t = \emptyset$ so that $\bar{P}_c^t = \bar{P}_c^0$:

$$\begin{aligned}
\text{P} [m_c \mid D^{\mathbf{g}_k^t}] &= \left(1 + \sum_{q=1}^{|\mathbf{g}_k^t|} (-1)^q \sum_{\boldsymbol{\eta} \subseteq \mathbf{g}_k^t, \boldsymbol{\eta} \in \mathbb{N}^q} \prod_{i=1}^q (1 - P_c^{\eta_i}) f_c^0(\boldsymbol{\eta}) \right) \bar{P}_c^0 \\
\text{P} [\bar{m}_c \mid D^{\mathbf{g}_k^t}] &= \left(1 + \sum_{q=1}^{|\mathbf{g}_k^t|} (-1)^q \sum_{\boldsymbol{\eta} \subseteq \mathbf{g}_k^t, \boldsymbol{\eta} \in \mathbb{N}^q} f_c^0(\boldsymbol{\eta}) \right) (1 - \bar{P}_c^0) ,
\end{aligned} \tag{3.14}$$

Substitution of these expressions into the last line of (3.13) yields

$$r_c^t = \prod_{k=1}^{|\mathcal{G}^t|} \frac{1 + \sum_{q=1}^{|\mathbf{g}_k^t|} (-1)^q \sum_{\boldsymbol{\eta} \subseteq \mathbf{g}_k^t, \boldsymbol{\eta} \in \mathbb{N}^q} \prod_{i=1}^q (1 - P_c^{\eta_i}) f_c^0(\boldsymbol{\eta})}{1 + \sum_{q=1}^{|\mathbf{g}_k^t|} (-1)^q \sum_{\boldsymbol{\eta} \subseteq \mathbf{g}_k^t, \boldsymbol{\eta} \in \mathbb{N}^q} f_c^0(\boldsymbol{\eta})} \cdot \bar{\rho}_c^t . \tag{3.15}$$

The OG mapping algorithm with (3.15) at its core is implemented in pseudocode as Algorithm 4.

Algorithm 4 The extended-standard OG mapping algorithm for approximating the posterior odds ratios. The required initializations are $\bar{\rho}_c^0 = \frac{P_c^0}{1-P_c^0}$, and $\rho_c^0 = 1$. The additional ratio $\rho_{c_F}^t$, which is analogous to ρ_c^t for all $\{d^t : \tau \in \bigcup_{k \in \mathcal{G}_F^t} \mathcal{G}_k^t\}$, is maintained to avoid permanent storage of the *ppt* for static subsets of detection times. If $|\mathcal{G}_k^t| = 1 \forall k$, this algorithm is equivalent to the standard algorithm for binary measurements (Algorithm 3).

Require:

- $$\begin{aligned} & \rho_c^{t-1}, \rho_{c_F}^{t-1}, \bar{\rho}_c^{t-1} \\ & P_c^t, P_F^t \\ & \mathcal{G}^t, \mathcal{G}_F^t, \mathcal{G}_F^{t-1} \\ & \mathbf{g}_i^t, \forall i \in \mathcal{G}^t \\ & P_c^{\tau_i}, \forall \tau_i \in \bigcup_{k \in \mathcal{G}^t \setminus \mathcal{G}_F^{t-1}} \mathcal{G}_k^t \end{aligned}$$
- 1: **if** d^t (detection) **then**
 - 2: $\bar{\rho}_c^t = \bar{\rho}_c^{t-1}$
 - 3: **if** $\mathcal{G}_F^t = \mathcal{G}_F^{t-1}$ **then**
 - 4: $\rho_{c_F}^t = \rho_{c_F}^{t-1}$
 - 5: **else**
 - 6: $k = \mathcal{G}_F^t \setminus \mathcal{G}_F^{t-1}$
 - 7:
$$\rho_{c_F}^t = \frac{1 + \sum_{q=1}^{|\mathcal{G}_k^t|} (-1)^q \sum_{\eta \subseteq \mathcal{G}_k^t, \eta \in \mathbb{N}^q} \prod_{i=1}^q (1 - P_c^{\eta_i}) f_c^q(\eta)}{1 + \sum_{q=1}^{|\mathcal{G}_k^t|} (-1)^q \sum_{\eta \subseteq \mathcal{G}_k^t, \eta \in \mathbb{N}^q} f_c^q(\eta)}$$
 - 8: **end if**
 - 9:
$$\rho_c^t = \rho_{c_F}^t \cdot \prod_{k \in \mathcal{G}^t \setminus \mathcal{G}_F^t} \frac{1 + \sum_{q=1}^{|\mathcal{G}_k^t|} (-1)^q \sum_{\eta \subseteq \mathcal{G}_k^t, \eta \in \mathbb{N}^q} \prod_{i=1}^q (1 - P_c^{\eta_i}) f_c^q(\eta)}{1 + \sum_{q=1}^{|\mathcal{G}_k^t|} (-1)^q \sum_{\eta \subseteq \mathcal{G}_k^t, \eta \in \mathbb{N}^q} f_c^q(\eta)}$$
 - 10: **else if** \bar{d}^t (nondetection) **then**
 - 11: $\bar{\rho}_c^t = (1 - P_c^t) \bar{\rho}_c^{t-1}$
 - 12: $\rho_c^t = \rho_c^{t-1}$
 - 13: **end if**
 - 14: $r_c^t = \rho_c^t \cdot \bar{\rho}_c^t$
-

If $|\mathcal{G}_k^t| = 1 \forall k$, then Algorithm 4 is identical to the standard algorithm. For larger subsets $\mathcal{G}_k^t \subseteq \tau^t$, Algorithm 4 retains the standard algorithm's property that detections and non-detections are incorporated into the map independently of one another. Thus, if a subset of detection times becomes static, it can be permanently

incorporated into the posterior and the corresponding data cleared from memory. In this work, Algorithm 2 is used to define the \mathbf{g}_k^t , so that all \mathbf{g}_k^t such that $k \in \mathcal{G}_F^t$ can be eliminated from memory.

3.1.3 Conditional Independence of Detections Algorithm

The relaxed CIM assumption (3.7b) requires conditional independence to hold only between subsets of detections while retaining the conditioning on non-detections and leads to an alternative approximate algorithm that retains the dependence of detections on non-detections. The derivation proceeds along much the same lines as for the extended-standard algorithm above, but I include it here for completeness. Again starting from (3.1), this time apply (3.7b), followed by Bayes Rule:

$$\begin{aligned}
r_c^t &= \frac{\mathrm{P} [D^t \mid \bar{D}^t, m_c]}{\mathrm{P} [D^t \mid \bar{D}^t, \bar{m}_c]} \cdot \bar{\rho}_c^t \\
&= \prod_{k=1}^{|\mathcal{G}^t|} \frac{\mathrm{P} [D^{\mathbf{g}_k^t} \mid \bar{D}^t, m_c]}{\mathrm{P} [D^{\mathbf{g}_k^t} \mid \bar{D}^t, \bar{m}_c]} \cdot \bar{\rho}_c^t \\
&= \prod_{k=1}^{|\mathcal{G}^t|} \frac{\mathrm{P} [m_c \mid D^{\mathbf{g}_k^t}, \bar{D}^t] \mathrm{P} [\bar{m}_c \mid \bar{D}^t]}{\mathrm{P} [\bar{m}_c \mid D^{\mathbf{g}_k^t}, \bar{D}^t] \mathrm{P} [m_c \mid \bar{D}^t]} \cdot \bar{\rho}_c^t.
\end{aligned} \tag{3.16}$$

To evaluate the arguments of the product above, again use the expressions for $\mathrm{P} [m_c \mid D^t, \bar{D}^t]$ and $\mathrm{P} [\bar{m}_c \mid D^t, \bar{D}^t]$ given by equations (B.29) and (B.30), only this time with the modification $D^t = D^{\mathbf{g}_k^t}$. Substitution of the resultant expressions into the last line of (3.16) yields

$$r_c^t = \prod_{k=1}^{|\mathcal{G}^t|} \frac{1 + \sum_{q=1}^{|\mathbf{g}_k^t|} (-1)^q \sum_{\boldsymbol{\eta} \subseteq \mathbf{g}_k^t, \boldsymbol{\eta} \in \mathbb{N}^q} \prod_{i=1}^q (1 - P_c^{\eta_i}) f_c^t(\boldsymbol{\eta})}{1 + \sum_{q=1}^{|\mathbf{g}_k^t|} (-1)^q \sum_{\boldsymbol{\eta} \subseteq \mathbf{g}_k^t, \boldsymbol{\eta} \in \mathbb{N}^q} f_c^t(\boldsymbol{\eta})} \cdot \bar{\rho}_c^t. \tag{3.17}$$

The OG mapping algorithm with (3.17) at its core is implemented in pseudocode as Algorithm 5.

The algorithms introduced previously do not compute exact posteriors after even a single detection. By retaining the conditioning on \bar{D}^t , this algorithm produces exact

Algorithm 5 The CID algorithm for approximating the odds ratio. The required initialization is $\bar{\rho}_c^0 = \frac{\bar{P}_c^0}{1 - \bar{P}_c^0}$.

Require:

$\bar{\rho}_c^{t-1}$
 P_c^t
 $P_c^{\tau_i}, \forall \tau_i \in \tau^{t-1}$
 \mathcal{G}^t
 $\mathbf{g}_i^t, \forall i \in \mathcal{G}^t$
1: **for all** $s \in \{1, \dots, C\}$ **do**
2: **if** d^t (detection) **then**
3: $\bar{\rho}_s^t = \bar{\rho}_s^{t-1}$
4: **else if** \bar{d}^t (nondetection) **then**
5: $\bar{\rho}_s^t = (1 - P_s^t) \bar{\rho}_s^{t-1}$
6: **end if**
7: $\bar{P}_s^t = \bar{\rho}_s^t / (1 + \bar{\rho}_s^t)$
8: **end for**
9: $\rho_c^t = \prod_{k=1}^{|\mathcal{G}^t|} \frac{1 + \sum_{q=1}^{|\mathbf{g}_k^t|} (-1)^q \sum_{\boldsymbol{\eta} \subseteq \mathbf{g}_k^t, \boldsymbol{\eta} \in \mathbb{N}^q} \prod_{i=1}^q (1 - P_c^{\eta_i}) f_c^q(\boldsymbol{\eta})}{1 + \sum_{q=1}^{|\mathbf{g}_k^t|} (-1)^q \sum_{\boldsymbol{\eta} \subseteq \mathbf{g}_k^t, \boldsymbol{\eta} \in \mathbb{N}^q} f_c^q(\boldsymbol{\eta})}$
10: $r_c^t = \rho_c^t \cdot \bar{\rho}_c^t$

results in regions of the map observed exclusively by detections belonging to the same subset of detection times. However, this concession necessitates batch processing of all detections following each measurement, including after non-detections because the marginal posteriors conditioned on exclusively non-detections \bar{P}_c^t enter into the $f_c^t(\boldsymbol{\eta})$. Unlike Algorithm 4 (and Algorithm 3, a special case of Algorithm 4), the P_c^t for every detection must therefore be permanently stored in memory. This algorithm is similar in operation to Algorithm 4 with the exception that the exact revised priors \bar{P}_c^t are used in place of the original priors. Since $\bar{P}_c^t \leq \bar{P}_c^0$, the CIM-like assumption (3.7b) is effectively required to hold in an environment with a reduced prior. Simulations presented later show that, especially when combined with an already low prior, this property can lead to inflated posteriors following detections like those generated by the standard algorithm.

3.2 Approximate Algorithms from Assuming Independence of the Posteriors

Section § 2.3.2 showed that the marginal posteriors conditioned exclusively on measurement times that resulted in non-detections $P[m_c \mid \bar{D}^t]$ remained independent of one another. The same property does not hold for posteriors when the conditioning includes detections because detections create dependencies between the observed cells. However, ignoring that fact and assuming that such conditional independence is maintained leads to an alternate class of algorithms that eliminate or further relax the modified CIM assumptions of the preceding section.

Though an assumption on the posteriors may seem initially unattractive since it explicitly violates the dependence that arises between them, it is perhaps natural in the context of occupancy grid mapping where the full posterior is unattainable except under the condition that the single-cell posteriors remain independent. Furthermore, algorithms assuming independence of the posteriors avoid requiring that measurements be interpreted completely independently. Though the dependencies between cells are not maintained, their “revised priors” reflect the belief accrued from previous measurements.

The algorithms developed below in § 3.2.1 all rely on assuming the single-cell marginal posteriors are independent conditioned on certain sets of detection times $\tau \in \tau^t$:

$$p(\mathbf{m} \mid \Delta^\tau, \Delta^{\bar{\tau}^\tau}) = \prod_{s=1}^C p(\mu_c \mid \Delta^\tau, \Delta^{\bar{\tau}^\tau}) . \quad (3.18)$$

If (3.18) holds for all detection times, then a recursive algorithm with memory requirements identical to the standard algorithm is attained. Alternately, the exact odds (2.21) can be exploited to produce a hybrid algorithm that requires an additional assumption akin to (3.7b) to be valid for the set of detection times $\{\mathbf{g}_k^t : k \in \mathcal{G}^t \setminus \mathcal{G}_F^t\}$.

The algorithms based on (3.18) essentially “revise the prior” to reflect the current posterior before incorporating a new measurement. This terminology reflects the shared property of independence between the marginal priors \bar{P}_c^0 and the true marginal posteriors when conditioned on non-detections alone \bar{P}_c^t . Inherent in this procedure is a dependence on the order with which measurements are incorporated into the map. The “revised prior” provides a weak linkage between previous mea-

measurements that resulted in detections and more recent ones; however, the linkage is uni-directional and these algorithms lack a mechanism to reinterpret previous measurements in light of the new data. Since these algorithms rely partially on the current state of the map to interpret new measurements, they are vulnerable to bias introduced by reinforcing the inaccuracies of a partially resolved map.

3.2.1 IP Algorithm

Assumptions of the form (3.18) lead to a novel class of OG mapping algorithms I call the Independence of Posteriors (IP) algorithms. In principle, independence of the posteriors could be assumed in conjunction with any sensor model; however, forward model (2.11) is special because the priors \bar{P}_c^0 enter explicitly into the inverse sensor model generated from it, thereby providing a mechanism for interpreting the current measurement in terms of the present belief in the states of the cells in the map. Unlike the related “context-sensitive” OG methods of [74,86], which also consider the present state of the map to interpret new measurements, the IP algorithms are applicable to any sensor satisfying (2.11) rather than specific to sonar range finders.

This section begins with the derivation of a recursive OG mapping algorithm with identical memory and processing requirements as the standard algorithm. The derivation is followed by the extensions required to capitalize on (2.21) for the preprocessing of subsets of detections. A discussion of the computational features and a pseudocode implementation of each algorithm follows.

Recursive IP Algorithm

Assumption (3.18) applied to all times $t \in \tau^t$ is sufficient to derive a recursive algorithm with memory and processing requirements identical to the standard algorithm. The derivation proceeds inductively. Suppose at time $t = \tau_n$ a detection is registered. Under the assumptions that $r_c^{\tau_n}$ is exact for all c and that $p(\mathbf{m} \mid \Delta^{\tau_n}) =$

$\prod_{s=1}^C \mathbb{P}(\mu_c \mid \Delta^{\tau_n})$, it can be shown that

$$r_c^t = (1 - P_c^t) r_c^{t-1}, \quad \tau_n < t < \tau_{n+1} \quad (3.19a)$$

$$r_c^{\tau_{n+1}} = \frac{1 - (1 - P_F^{\tau_{n+1}})(1 - P_c^{\tau_{n+1}}) \prod_{s \neq c} (1 - P_s^{\tau_{n+1}} \tilde{P}_s^{\tau_{n+1}})}{1 - (1 - P_F^{\tau_{n+1}}) \prod_{s \neq c} (1 - P_s^{\tau_{n+1}} \tilde{P}_s^{\tau_{n+1}})} \cdot r_c^{\tau_{n+1}-1} \quad (3.19b)$$

where τ_{n+1} denotes the next time $t > \tau_n$ at which a detection occurs and the \tilde{P}_c^τ denote the approximate revised priors. At $t = \tau_{n+1} - 1$, these can be computed from $r_c^{\tau_{n+1}-1}$ as

$$\tilde{P}_c^{\tau_{n+1}} = \frac{r_c^{\tau_{n+1}-1}}{(1 + r_c^{\tau_{n+1}-1})}. \quad (3.20)$$

Since $r_c^{\tau_1}$ computed using (3.19a & b) is exact by (2.21), it follows by induction that r_c^t computed this way is exact for all t if at each $\tau \in \boldsymbol{\tau}^t$ independence of the posteriors is assumed:

$$\mathbb{P}(\boldsymbol{m} \mid \Delta^\tau) = \prod_{s=1}^C \mathbb{P}(\mu_c \mid \Delta^\tau), \quad \tau \in \{\boldsymbol{\tau}^t\}. \quad (3.21)$$

The remaining step of showing (3.19a & b) for $\tau_n < t < \tau_{n+1}$ is given in App. D.1.

The algorithm defined by (3.19) has the same memory and computation requirements as the standard algorithm; however, it behaves quite differently. The unidirectional linkage of past detections to new ones via the “revised prior” is demonstrated in Fig. 3-1. An implementation in pseudocode is provided as Algorithm 6.

Extended IP Algorithm

Extending the recursive form of the IP algorithm for $|\mathbf{g}_k^t| \neq 1$ is straightforward, but requires the additional modified CIM assumption on all subsets of detection times $\{\mathbf{g}_k^t : k \in \mathcal{G}^t \setminus \mathcal{G}_F^t\}$ given below. For the purposes of the following derivation, it is convenient to define the “revised prior” odds, $\tilde{\rho}_c^t \triangleq \tilde{P}_c^t / (1 - \tilde{P}_c^t)$ and the set of measurement times that are members of fixed subsets, $\boldsymbol{\tau}_F^t \triangleq \bigcup_{k \in \mathcal{G}_F^t} \mathbf{g}_k^t$.

As above, the derivation proceeds inductively. Suppose at time $t = \tau_n$ a detection is registered and that $\mathcal{G}_F^{\tau_n} \supset \mathcal{G}_F^{\tau_n-1}$ where \supset denotes a strict superset. That is, at time

	Standard Alg.	Revised Prior Alg.	Exact Alg.
$t = 0$: prior	<div> <div>1 0.010</div> <div>2 0.010</div> <div>3 0.010</div> <div>4 0.010</div> </div>	<div> <div>1 0.010</div> <div>2 0.010</div> <div>3 0.010</div> <div>4 0.010</div> </div>	<div> <div>1 0.010</div> <div>2 0.010</div> <div>3 0.010</div> <div>4 0.010</div> </div>
$t = 1$: detection $D^1 = \{d^1\}$, $\bar{D}^1 = \emptyset$ $\mathcal{R}^1 = \{1, 2, 3\}$	<div> <div>1 0.340</div> <div>2 0.340</div> <div>3 0.340</div> <div>4 0.010</div> </div>	<div> <div>1 0.340</div> <div>2 0.340</div> <div>3 0.340</div> <div>4 0.010</div> </div>	<div> <div>1 0.340</div> <div>2 0.340</div> <div>3 0.340</div> <div>4 0.010</div> </div>
$t = 2$: detection $D^2 = \{d^1, d^2\}$, $\bar{D}^2 = \emptyset$ $\mathcal{R}^1 = \{2, 3, 4\}$	<div> <div>1 0.340</div> <div>2 0.963</div> <div>3 0.963</div> <div>4 0.340</div> </div>	<div> <div>1 0.340</div> <div>2 0.659</div> <div>3 0.659</div> <div>4 0.024</div> </div>	<div> <div>1 0.023</div> <div>2 0.506</div> <div>3 0.506</div> <div>4 0.023</div> </div>

Figure 3-1: Recursive update of a 4-cell map via the standard and recursive IP algorithms versus the exact result. The sensor characteristics were $P_c^t \equiv 0.1$ for all cells within each detection region and $P_F^t \equiv 0$. All three algorithms produce the identical result following the first detection. Following the second detection, neither the standard nor the IP algorithms update the posterior in cell 1; however, the standard algorithm also grossly over-estimates the posteriors in cells 2–4. This is because the detection at time $t = 2$ is interpreted without enforcing consistency with the previous detection or with the prior. The IP algorithm only slightly raises the posterior in cell 4 because the high revised prior of cells 2 and 3 following the detection at $t = 1$ make the detection at $t = 2$ highly likely to have originated from one of these cells. The posteriors in cells 2 and 3 are slightly higher than correct, a consequence of regarding the map at $t = 1$ as describing a new prior and effectively making a map with two occupied cells more likely than suggested by the original prior.

Algorithm 6 The recursive version of the IP algorithm for approximating the odds ratio. The required initialization is $r_c^0 = \bar{P}_c^0 / (1 - \bar{P}_c^0)$.

Require:

- r_c^{t-1}
 P_c^t
 $t > 0$
 1: **if** d^t (detection) **then**
 2: $\tilde{P}_c^t = \frac{r_c^{t-1}}{1+r_c^{t-1}}$
 3: $r_c^t = \frac{1 - (1 - P_F^t)(1 - P_c^t) \prod_{s \neq c} (1 - P_s^t \tilde{P}_s^t)}{1 - (1 - P_F^t) \prod_{s \neq c} (1 - P_s^t \tilde{P}_s^t)} \cdot r_c^{t-1}$
 4: **else if** \bar{d}^t (nondetection) **then**
 5: $r_c^t = (1 - P_c^t) r_c^{t-1}$
 6: **end if**
-

$t = \tau_n$ the subset of detection times to which τ_n was added is now immutable. Under the assumption that $\tilde{\rho}_c^{\tau_n-1}$ is exact and that the posteriors are independent

$$p(\mathbf{m} \mid \Delta^{\tau_F}, \Delta^{\bar{\tau}^{\tau_n}}) = \prod_{s=1}^C p(\mu_c \mid \Delta^{\tau_F}, \Delta^{\bar{\tau}^{\tau_n}}),$$

the “revised prior odds” $\tilde{\rho}_c^{\tau_n}$ can be computed readily. Let $\mathbf{g}_F = \{\mathbf{g}_k^{\tau_n} : \tau_n \in \mathbf{g}_k^{\tau_n}\}$ denote the subset of detection times to which the detection at $t = \tau_n$ was added. It can then be shown that

$$\tilde{\rho}_c^{\tau_n} = \tilde{\rho}_c^{\tau_n-1} \cdot \frac{1 + \sum_{q=1}^{|\mathbf{g}_F|} (-1)^q \sum_{\boldsymbol{\eta} \subseteq \mathbf{g}_F, \boldsymbol{\eta} \in \mathbb{N}^q} \prod_{i=1}^q (1 - P_c^{\eta_i}) \tilde{f}_c(\tau_n)}{1 + \sum_{q=1}^{|\mathbf{g}_F|} (-1)^q \sum_{\boldsymbol{\eta} \subseteq \mathbf{g}_F, \boldsymbol{\eta} \in \mathbb{N}^q} \tilde{f}_c(\tau_n)}, \quad (3.22)$$

where

$$\tilde{f}_c^{\tau_n}(\boldsymbol{\eta}) = \prod_{i=1}^{|\boldsymbol{\eta}|} (1 - P_F^{\eta_i}) \prod_{s \neq c} \left(1 - \left(1 - \prod_{i=1}^{|\boldsymbol{\eta}|} (1 - P_s^{\eta_i}) \right) \tilde{P}_s^{\tau_n} \right). \quad (3.23)$$

Under the additional modified CIM assumption given by

$$p(\Delta^{\tau^t \setminus \tau_F^t} \mid \Delta^{\tau_F^t}, \Delta^{\bar{\tau}^t}, \mu_c) = \prod_{k \in \mathcal{G}^t \setminus \mathcal{G}_F^t} p(\Delta^{\mathbf{g}_k^t} \mid \Delta^{\tau_F^t}, \Delta^{\bar{\tau}^t}, \mu_c) \quad (3.24)$$

it follows that

$$\tilde{\rho}_c^t = \begin{cases} (1 - P_c^t) \tilde{\rho}_c^{t-1}, & \{t: \mathcal{G}_F^t = \mathcal{G}_F^{\tau_n}, t \notin \tau^t\} \\ \tilde{\rho}_c^{t-1}, & \{t: \mathcal{G}_F^t = \mathcal{G}_F^{\tau_n}, t \in \tau^t\} \end{cases} \quad (3.25a)$$

$$r_c^t = \prod_{k \in \mathcal{G}^t \setminus \mathcal{G}_F^t} \frac{1 + \sum_{q=1}^{|\mathbf{g}_k^t|} (-1)^q \sum_{\boldsymbol{\eta} \subseteq \mathbf{g}_k^t, \boldsymbol{\eta} \in \mathbb{N}^q} \prod_{i=1}^q (1 - P_c^{\eta_i}) \tilde{f}_c(t)}{1 + \sum_{q=1}^{|\mathbf{g}_k^t|} (-1)^q \sum_{\boldsymbol{\eta} \subseteq \mathbf{g}_k^t, \boldsymbol{\eta} \in \mathbb{N}^q} \tilde{f}_c(t)} \cdot \tilde{\rho}_c^t, \quad \{t: \mathcal{G}_F^t = \mathcal{G}_F^{\tau_n}\}. \quad (3.25b)$$

The inductive base case corresponds to all $t > 0$ such that $\mathcal{G}_F^t = \emptyset$. In the base case, the above is identical to the CID algorithm (§ 3.1.3), and assumption (3.24) is identical to the modified CIM assumption required by the CID algorithm. Therefore under assumption (3.24), the above produces exact results in the base case. It follows by induction that r_c^t computed via (3.25a)—(3.25b) is exact if in addition to (3.24), conditional independence of the posteriors is assumed according to:

$$p(\mathbf{m} \mid \Delta^{\tau_F}, \Delta^{\bar{\tau}^t}) = \prod_{s=1}^C p(\mu_c \mid \Delta^{\tau_F}, \Delta^{\bar{\tau}^t}) , \quad t \in \{t: \mathcal{G}_F^t \supset \mathcal{G}_F^{t-1}\}. \quad (3.26)$$

The remaining steps are provided in App. D.2.

This extended IP algorithm shares most of its properties with the recursive form, though the core assumption on the independence of the posteriors need only apply for a smaller set of detection times $t \in \{\tau: \mathcal{G}_F^\tau \supset \mathcal{G}_F^{\tau-1}\}$. The price is an increased memory requirement (all P_c^τ for $\tau \in \tau^t \setminus \tau_F^t$ must be stored) and of course the added CIM assumption given by (3.24).

Despite the need for a CIM assumption like that for the CID algorithm, there is a crucial difference between these algorithms. The present algorithm rarely requires making a CIM assumption on subsets of detection times that observe similar portions of the map. Instead it will usually revise the prior using a complete subset before instantiating a new subset in the same region of the map. Reasonable mechanisms for generating the subsets of detection times \mathbf{g}_k^t should fill the active subset in a portion of the map (add it to \mathcal{G}_F^t) before instantiating a new subset from a detection in the same region. That said, depending on the nature of the algorithm used to

Algorithm 7 The extended IP algorithm for approximating the odds ratio.

Require:

$\rho_c^{t-1}, \bar{\rho}_c^{t-1}$
 P_c^t
 $P_c^{\tau_i}, \forall \tau_i \in \tau^{t-1}$
 $\mathcal{G}^t, \mathcal{G}_F^t$
 $\mathbf{g}_i^t, \forall i \in \mathcal{G}^t$
1: **for all** $s \in \{1, \dots, C\}$ **do**
2: **if** d^t (detection) **then**
3: **if** $\mathcal{G}_F^t \neq \emptyset$ **then**
4: $k = \mathcal{G}_F^t$
5:
$$\bar{\rho}_s^t = \bar{\rho}_s^{t-1} \cdot \frac{1 + \sum_{q=1}^{|\mathbf{g}_k^t|} (-1)^q \sum_{\eta \subseteq \mathbf{g}_k^t, \eta \in \mathbb{N}^q} \prod_{i=1}^q (1 - P_c^{\eta_i}) f_{\mathcal{A}}^t(\eta)}{1 + \sum_{q=1}^{|\mathbf{g}_k^t|} (-1)^q \sum_{\eta \subseteq \mathbf{g}_k^t, \eta \in \mathbb{N}^q} f_{\mathcal{A}}^t(\eta)}$$

6: $\mathcal{G}^t = \mathcal{G}^t \setminus \mathcal{G}_F^t$
7: $\mathcal{G}^t = \emptyset$
8: **else**
9: $\bar{\rho}_s^t = \bar{\rho}_s^{t-1}$
10: **end if**
11: **else if** \bar{d}^t (nondetection) **then**
12: $\bar{\rho}_s^t = (1 - P_s^t) \bar{\rho}_s^{t-1}$
13: **end if**
14: $\bar{P}_s^t = \bar{\rho}_s^t / (1 + \bar{\rho}_s^t)$
15: **end for**
16:
$$\rho_c^t = \prod_{k=1}^{|\mathcal{G}^t|} \frac{1 + \sum_{q=1}^{|\mathbf{g}_k^t|} (-1)^q \sum_{\eta \subseteq \mathbf{g}_k^t, \eta \in \mathbb{N}^q} \prod_{i=1}^q (1 - P_c^{\eta_i}) f_{\mathcal{A}}^t(\eta)}{1 + \sum_{q=1}^{|\mathbf{g}_k^t|} (-1)^q \sum_{\eta \subseteq \mathbf{g}_k^t, \eta \in \mathbb{N}^q} f_{\mathcal{A}}^t(\eta)}$$

17: $r_c^t = \rho_c^t \cdot \bar{\rho}_c^t$

assign detection times to the \mathbf{g}_k^t , situations may occasionally arise where two initially disparate \mathbf{g}_k^t will end up overlapping strongly following additional detections before either is assigned to \mathcal{G}_F^t . Algorithm 7 implements (3.25a)—(3.25b) in pseudocode.

3.3 Simulation

This section presents two sets of numerical simulations that compare the behavior of the standard occupancy grid mapping algorithm and the three new approximate algorithms proposed in the previous section to each other, and when possible, to the exact marginal posterior odds ratio (2.21).

The first set of investigations examines a short, deterministic sequence of successive detections with no intervening non-detections. The results demonstrate the appearance of inflated odds as a consequence of the CIM assumption. The new approximate algorithms produce better estimates of the true posterior than the standard algorithm, but their relative quality is dependent on how individual measurements are grouped. Because only a small number of detections are considered, the exact posterior odds can be computed and the approximate algorithms compared directly.

Results follow from simulations of a long record of binary measurements produced from repeated observation of a sparsely occupied environment at randomly generated locations. The large number of measurements prohibits computation of the exact posterior odds, and alternative map evaluation metrics are explored to facilitate inter-comparison of the results produced by the approximate algorithms.

A major difference between these simulations and real surveys conducted by autonomous vehicles is the random sampling pattern used here. The difference is deliberate and reflects the expectation that sampling trajectories, e.g. along regular track-lines, will interact with the approximations required by each algorithm. However, there are important distinctions in the behaviors of the various algorithms studied that become apparent without the additional complication of trackline artifacts.

Simulation Environment

The simulation environment consists of a square two-dimensional grid with a uniform prior probability of occupancy $P[m_c]$. The forward sensor model conforms to the form of (2.11), with the single-source detection probabilities P_c^t specified according to¹

$$P_c^t = P_{D_{max}} e^{-\frac{1}{2\sigma^2} \|\mathbf{x}^t - \mathbf{x}_c\|^2}, \quad (3.27)$$

¹To speed calculation, values of $P_c^t < 0.001$ were approximated as zero.

where \mathbf{x}^t and \mathbf{x}_c denote the sensor location and the location of node c respectively.

Table 3.1 lists numerical values of all constants used in the simulations. The choices reflect the nature of the chemical plume tracing problem [30]: relatively few sources, a measurement record composed mostly of non-detections, a high degree of spatial ambiguity between detection location and source location, and a low probability of false alarm.

Table 3.1: Parameter values used in simulations.

Environment		
C	1600 or 10000	Number of grid nodes
$P[m_c]$	0.001	Prior probability of occupancy
Forward Model		
$P_{D_{max}}$	0.4	Maximum single-source detection probability
σ	3 or 6	Sensor aperture length scale
P_F^t	0 or 0.01	Probability of false detection (constant)
Approximate Algorithms		
GROUPING_THRESHOLD	0.001	
MAX_GROUP_MEMBERS	varied	

Algorithm Demonstration, Detections Only

Figures 3-2 and 3-3 show maps generated from a deterministic sequence of five (Fig. 3-2) and ten (Fig. 3-3) detections proceeding counter-clockwise in a ring about the center of the map. Relative to the correct posterior odd ratios, the standard algorithm produces a map with very high odds across a broad peak at the center of the map. The ring-shaped ridge of the correct map is wholly absent.

In contrast, both the CID and IP approximate algorithms produce maps that to varying degrees preserve the ring-like structure of the correct posterior odds and to a lesser extent the actual value of the posterior odds. The variability within these maps reflects different partitions of τ^t and the different mechanisms by which the two algorithms process groups of maximal size. The largest deviations from the true posterior odds occur where the \mathbf{g}_k^t overlap strongly. Comparison of the CID results with sequentially composed subsets of 3 and 5 indicates improved performance for

larger subset size; however the major artifact in the $|\mathbf{g}_k^t| \leq 3$ result is a manifestation of unequal subset size (three groups of 3 plus one group of 1). The significantly degraded map produced by randomly assigning measurements to subsets illustrates the dependence of these results on how detections are grouped. Sequential grouping produces good results in this case because nearby measurements are grouped together. The same arguments hold for the maps generated by the IP algorithm, though the results are superior to those generated by the CID algorithm. Since all measurements are detections, the extended-standard algorithm produces results identical to the CID algorithm.

Algorithm Demonstration, Simulated Measurements

Figures 3-4, 3-5, & 3-6 show occupancy grid maps generated in simulation by the approximate algorithms proposed above along with the standard algorithm. For these simulations, a random groundtruth map of occupied cells was generated by sampling from the binary prior to assign the state of each node. Measurement locations were generated by sampling from a uniform distribution over the spatial domain of the grid. The binary sensor measurements were generated by sampling from the binary forward model (2.11) with the P_c^t specified according to the position of each measurement location relative to all occupied cells in the map via (3.27). Insights into the behaviors and performance of each algorithm are facilitated by viewing the simulation results at several time steps.

$t = 200$, **Figure 3-4**

The maps produced by each algorithm after 200 measurements (Fig. 3-4) already show peaks near many of the occupied cells and indicate confident emptiness over large areas. However, there are significant differences between the maps. With the exception of the IP results, the maps using no grouping of detections into dependent subsets ($|\mathbf{g}_k^t| = 1$) already show evidence of inflated posteriors at $t = 200$ in regions of dense detections. For instance, the main peak of the standard result contains roughly 100 cells whose posterior probability of occupancy is at least 0.10, suggesting some 10 cells of the 100 are occupied, when in fact only two are occupied. The CID show even more inflated posteriors. The algorithms produced using $|\mathbf{g}_k^t| \leq 6$ show less evidence

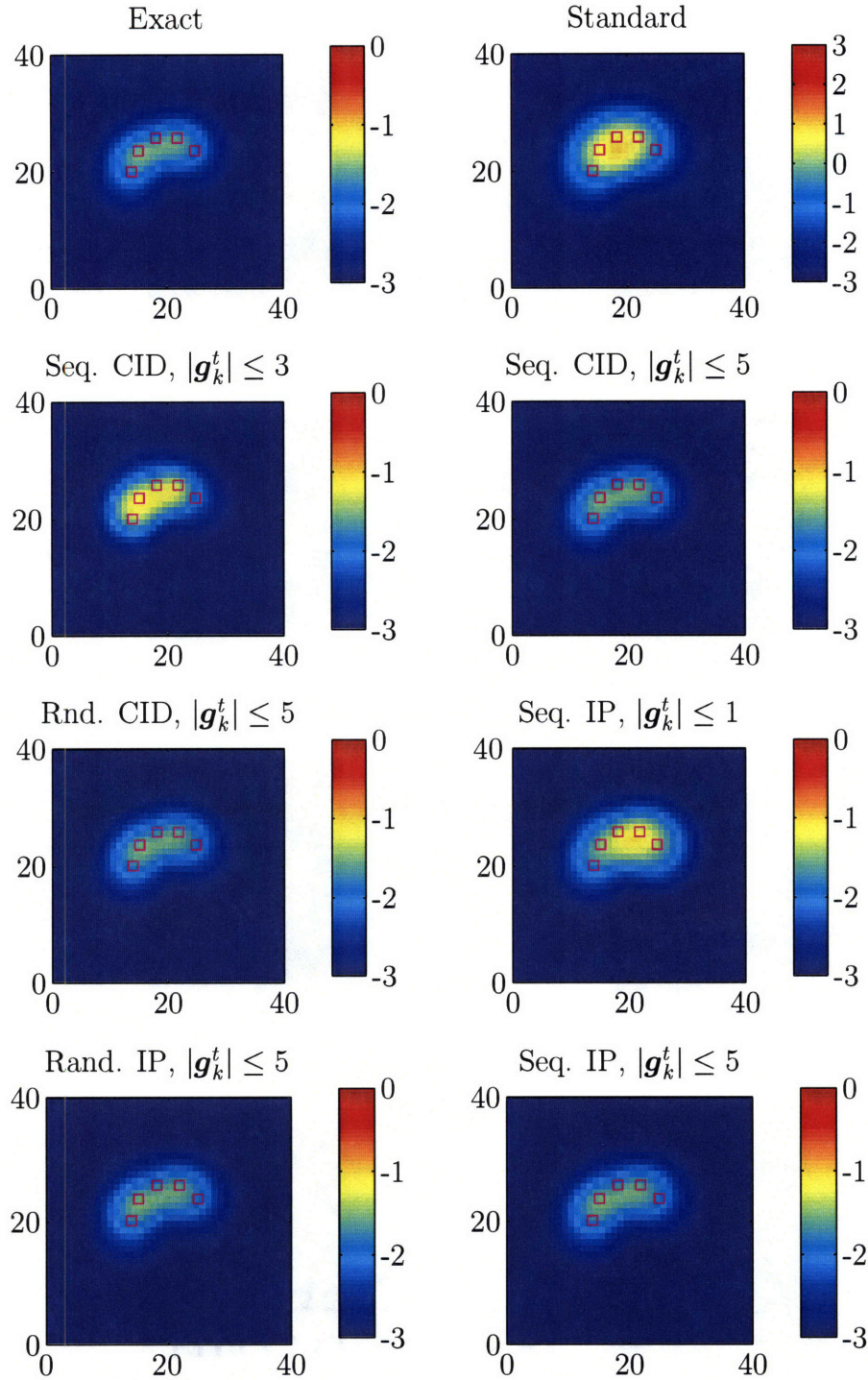


Figure 3-2: Occupancy grid maps after $n^t = 5$ successive detections computed exactly (upper left), via the standard algorithm (upper right) and via the new approximate algorithms for different group sizes and grouping schemes (remaining panels). The color of each cell is scaled according to $\log_{10} r_c^t$, with hot colors indicating a high posterior odds of occupancy. The CID and IP results for groups of five measurements are identical to the exact solution. The color scale spans a larger range for the standard algorithm to avoid clipping.

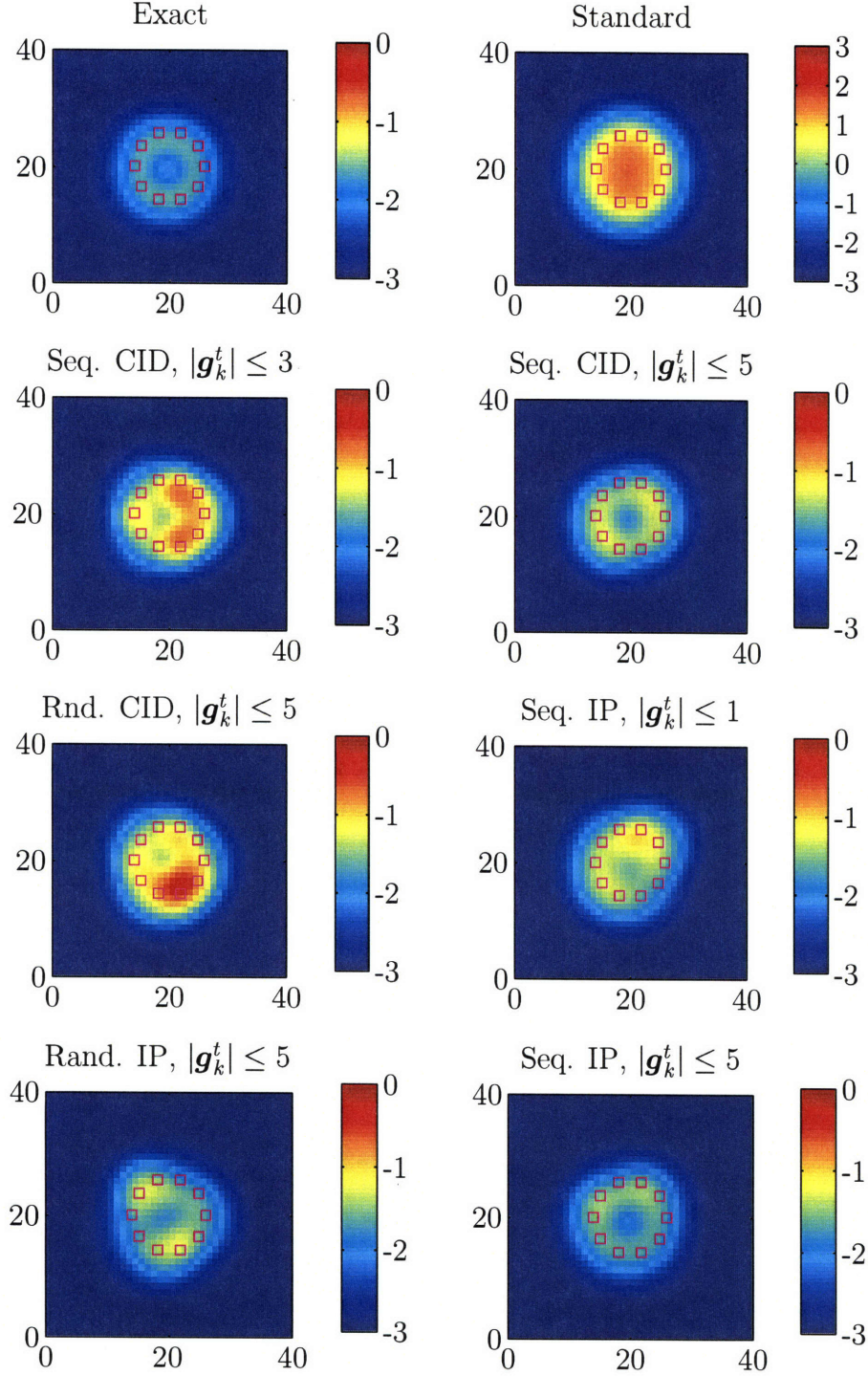


Figure 3-3: The occupancy grid maps of Fig. 3-2, but after $n^t = 10$ successive detections. The influence of different groupings of measurements is evident in the results for the three instances each of the CID and IP algorithms; however, all instances outperform the standard algorithm in terms of the accuracy of the estimated posteriors relative to the exact result.

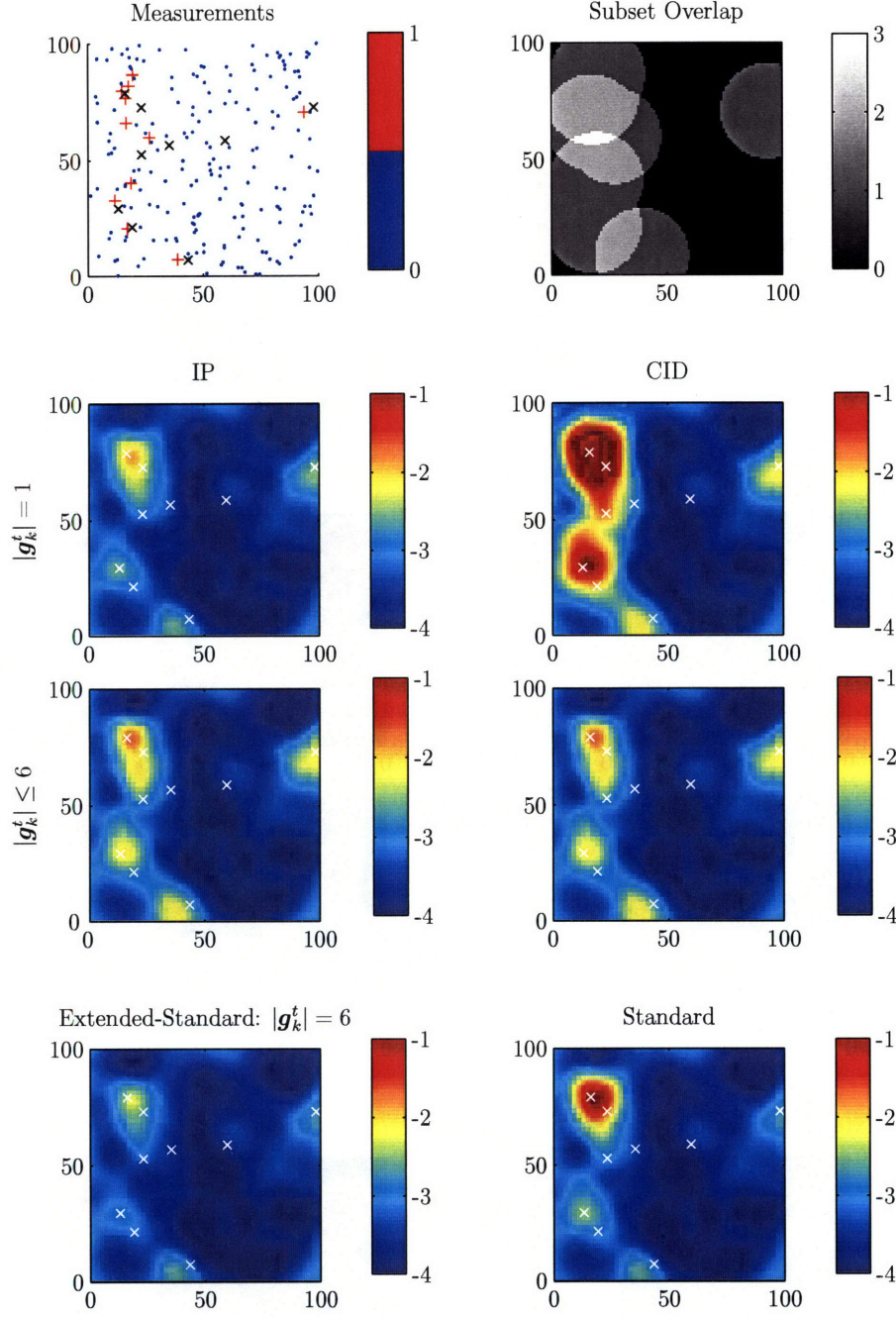


Figure 3-4: Occupancy grid maps generated in simulation by four approximate OG mapping algorithms, including the standard algorithm. The upper plots of the IP and CID results were produced with $|g_k^t| = 1$; the lower plots with $|g_k^t| \leq 6$. In each map occupied cells are marked by white x's. These plots show the maps after 200 measurements, most of which were non-detections. The plot in the upper left shows the spatial location of each measurement and its result: detection (red), or non-detection (blue). The grayscale plot applies to the results for $|g_k^t| \leq 6$ only. The color of each cell indicates the number of subsets of detections that include at least one observation of that cell. Posteriors within regions of the maps produced by the CID and IP algorithms with $|g_k^t| \leq 6$ observed exclusively by detections belonging to the same subset are exact.

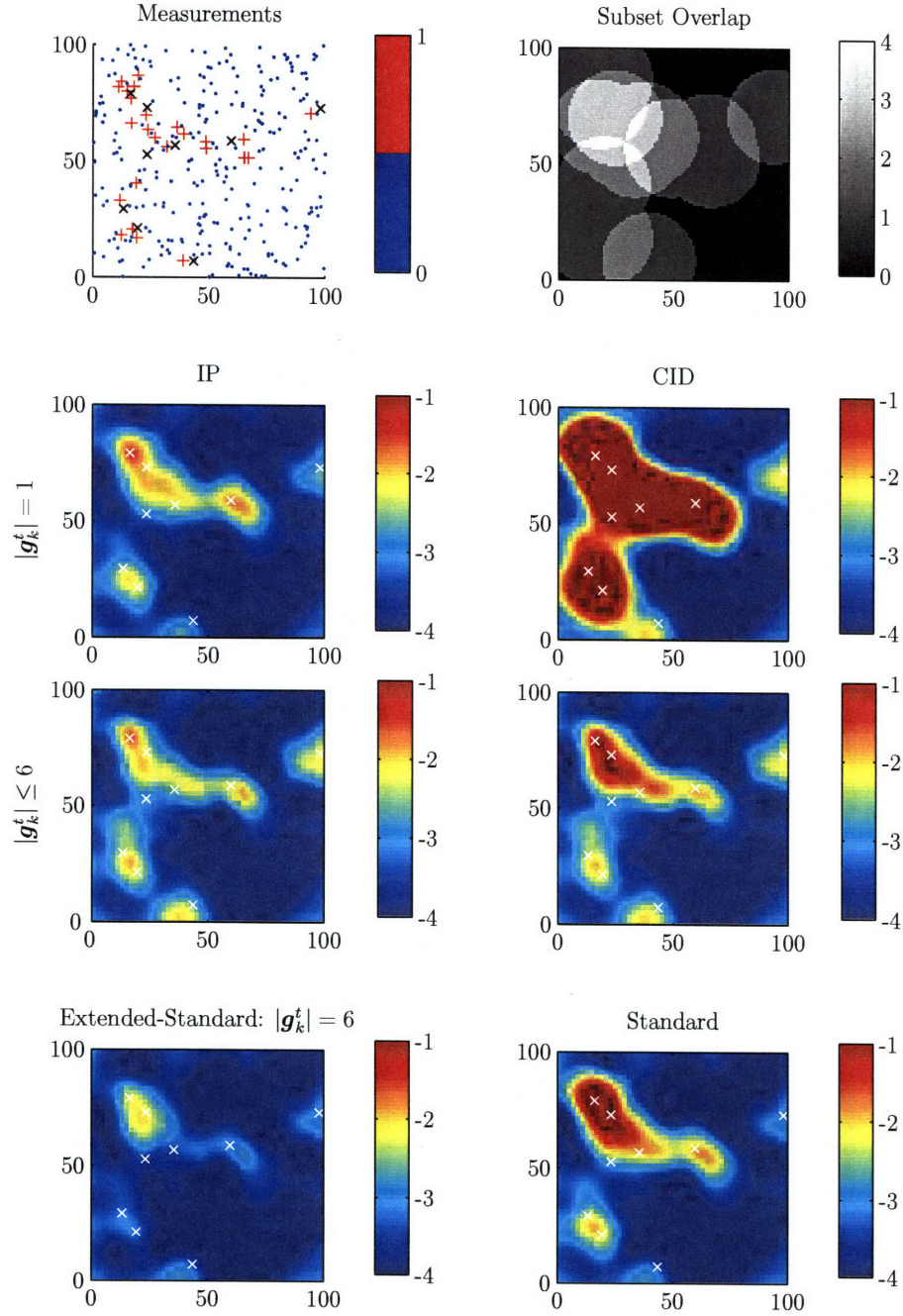


Figure 3-5: The maps of Fig. 3-4 after 300 measurements.

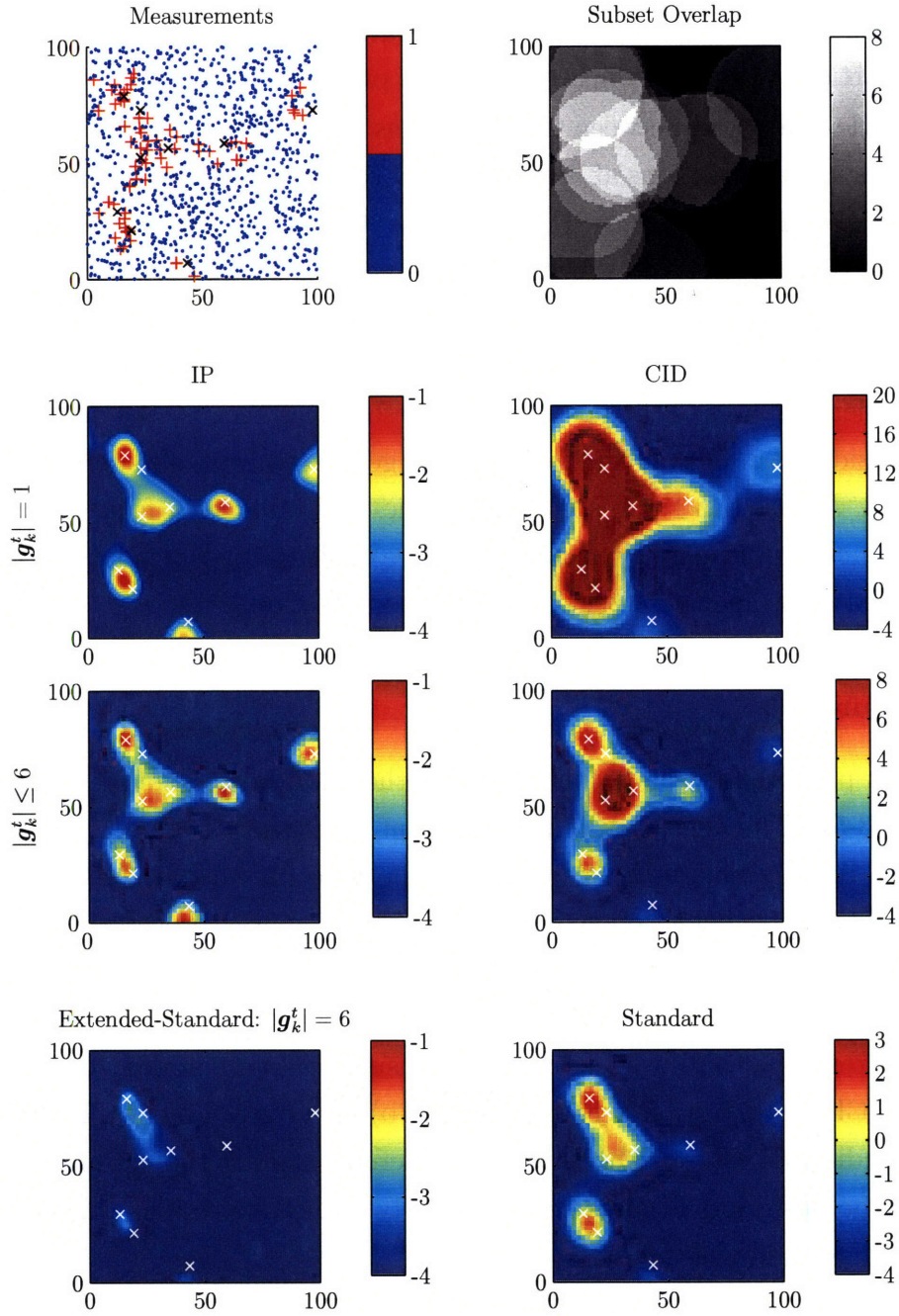


Figure 3-6: The maps of Figs. 3-4 & 3-5 after 1000 measurements. The axes have been rescaled to facilitate interpretation of the CID and standard results.

of inflated posteriors, a consequence of the less stringent CIM assumptions required. Additionally, the peaks are more uniform, reflecting the strong correlation between nearby detections. The extended-standard results appear the least satisfying, with almost no portions of the map raised above the prior.

More subtle differences between the algorithms are manifest near the occupied cells observed by isolated detections at approximately (45, 10) and (95, 70). In the $|\mathbf{g}_k^t| \leq 6$ maps produced using the IP and CID algorithms, these posteriors are relatively high. Since all $|\mathbf{g}_k^t| \leq 6$ at the time shown, the IP and CID results are in fact identical. With no other nearby detections, the posteriors near (95, 70) are exact, and nearly so around (45, 10). All algorithms using $|\mathbf{g}_k^t| = 1$ except the CID algorithm have produced underestimates of the posteriors in these places because the detections have not been considered in light of the revised prior generated from non-detections alone. With the low P_F^t used in these simulations, the possibility of a false alarm is very nearly nil, hence requiring overwhelming negative evidence to contradict the existence of an occupied cell somewhere near each isolated detection.

$t = 300$, **Figure 3-5**

For the most part, these trends continue with the addition of another 100 measurements (Fig. 3-5). The exception is the apparent divergence of the $|\mathbf{g}_k^t| \leq 6$ IP and CID results. Shortly before the time shown, the \mathbf{g}_k^t centered near (20, 80) achieved capacity. The nearby detection momentarily thereafter was processed as if it were independent by each algorithm, with significantly different results. In contrast to the map produced by the IP algorithm, the CID result shows an apparently inflated posterior, a trend that will worsen as more measurements are added. Unfortunately, relaxing the CIM assumption according to (3.7b) effectively lowers the prior before requiring conditional independence to hold, thereby making the assumption worse. Eventually, the posteriors conditioned exclusively on non-detections will become low enough to cause numerical failure.

$t = 1000$, **Figure 3-6**

By $t = 1000$ (Fig. 3-6), only the extended-standard and IP results are free from obviously inflated posteriors. Most occupied cells in the extended-standard map are asso-

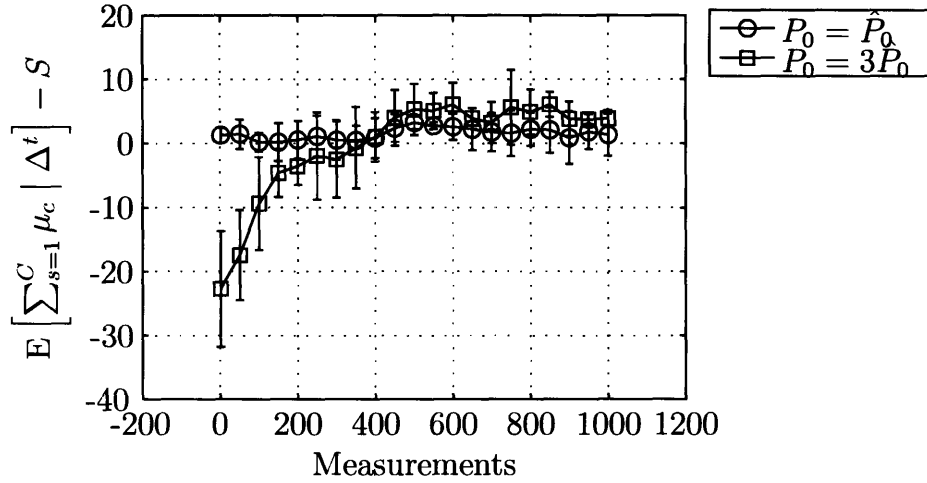


Figure 3-7: Error between the actual and expected number of occupied cells versus time for maps produced using the IP algorithm. Each curve represents the average of 3 trials using either the correct prior $\hat{P}_0 = P_0$ or else the underestimate $\hat{P}_0 = \frac{1}{3}P_0$. All trials were conducted with $|g_k^t| = 1$. The error bars indicate standard deviation. Results with $|g_k^t| \leq 6$ were within one standard deviation of the results shown.

ciated with posteriors below the prior making the mapping result essentially worthless. In contrast, the IP maps exhibit peaks of relatively uniform magnitude near all occupied cells (see also Fig. 3-9). Nevertheless, all maps except the extended-standard map in Fig. 3-6 show raised posteriors near occupied cells and could potentially be useful to guide a refined search. I explore this idea in the next section.

Summary

Most of the maps of Figs. 3-4, 3-5, & 3-6 indicate high posteriors over large areas that are grossly inconsistent with both the assumed low prior ($\bar{P}_c^0 \equiv 0.001$) and the actual number of occupied cells (nine). The IP results do not show obviously high posteriors. Indeed, Fig. 3-7 supports the contention that the algorithm is capable predicting the actual number of occupied cells even when that number does not correspond well to the assumed prior. The estimates shown in Fig. 3-7 are the expected number of occupied cells:² $E \left[\sum_{s=1}^C \mu_s \mid \Delta^t \right] = \sum_{s=1}^C P \left[m_s \mid \Delta^t \right]$.

Obviously these results apply strictly only to the parameters tested here. Empir-

²Higher order moments are problematic to estimate without assuming independence of the marginal posteriors.

ical evidence from addition simulations suggests these inter-algorithm trends persist over a range of parameters; however two qualitative observations are of note. (1) Increasing sensor aperture reduces the quality of the results, not only because the sensor is consequently less specific, but because measurements tend to be more correlated and cell posteriors more dependent thereby exacerbating the weaknesses of these algorithms. (2) Increasing P_F^t reduces the correlation between measurements and tends to blur the distinctions between the algorithms described here. Of course the degradation in sensor performance implied by high P_F^t also produces degraded mapping results.

Map Evaluation

The large number of detections in the simulations presented above prevent computation of the the correct posterior via (2.21) and hence preclude evaluation of these maps in terms of absolute error. Many previous investigations into OG mapping have relied on entropy to evaluate maps and to guide robotic exploration [79, 118]. The entropy of the posterior on μ_c is:

$$H_{\mu_c | \Delta^t} = -P[m_c | D^t, \bar{D}^t] \log_2(P[m_c | D^t, \bar{D}^t]) - (1 - P[m_c | D^t, \bar{D}^t]) \log_2(1 - P[m_c | D^t, \bar{D}^t]) \quad (3.28)$$

This expression captures the information content in the posterior. An entropy of 1 corresponds to total uncertainty ($P[m_c | D^t, \bar{D}^t] = 0.5$); and entropy of 0 corresponds to complete certainty in either occupancy ($P[m_c | D^t, \bar{D}^t] = 1$) or emptiness ($P[m_c | D^t, \bar{D}^t] = 0$). Assuming independence of the posteriors, the total entropy of an OG map is given by the sum $\sum_{s=1}^C H_{\mu_s | \Delta^t}$.³

Unfortunately, the utility of entropic measures of map quality are limited in the context of a low prior. Low priors imply a very low summed entropy before any measurements are made. Subsequent non-detections do reduce map entropy, and since independence of the posteriors is maintained when conditioned on exclusively non-detections, the summed entropy remains exact. In contrast, detections tend to raise

³Singh [109] offered another perspective on entropic measures in OG mapping. He investigated the *rate* of entropy change to evaluate the efficiency of sonar mapping with respect to sensing modality, navigation uncertainty, and survey parameters.

summed map entropy, and furthermore, because the posteriors are no longer independent, the summed entropy no longer represents a measure of the true information content of the map. This behavior persists even in the over-confidently occupied maps produced by some of the algorithms explored above. Though large portions of the maps produced by those algorithms are confidently occupied or empty, the boundary between these regions tends to increase in length as the area of the confidently occupied regions increases. The alternative measure of map quality introduced next is motivated by the intended use of these maps to guide subsequent higher-resolution survey in small areas around probable occupied cells.

Suppose that a fixed fractional percentage of the map likely to be occupied will be revisited (re-surveyed) on a subsequent survey to refine the locations of occupied cells within these regions. In that case, precise posteriors are not critical as long as the highest posteriors occur near occupied cells. From this perspective better maps will have more tightly constrained high posterior regions around each occupied cell. Figure 3-8 shows the fraction of the total number of occupied cells within the revisited portion of each map for various fixed revisitation percentages averaged over 10 trials. Prior to $t \approx 150$ all the algorithms produce, on average, nearly identical results. This is because early non-detections act to reduce the posterior in large, uninterrupted areas of unoccupied space. At times $t > 200$ the differences in scores between the algorithms reflect the different methods used by each to process detections. Only the IP algorithm shows consistently better performance for the revisitation-fractions investigated. The largest differences between using no grouping ($|\mathbf{g}_k^t| = 1$) and the largest maximum group size investigate ($|\mathbf{g}_k^t| \leq 6$) are apparent in the CID results (bottom panel, Fig. 3-8). With $|\mathbf{g}_k^t| \leq 6$, the decline in map score evident in the top panel of Fig. 3-8 after $t \approx 200$ is delayed. Nevertheless, even under this further relaxed CIM assumption, the CID results are superior to the standard algorithm results over a relatively short interval before severely inflated posteriors begin to degrade the result. There is also mild improvement in the IP results for $|\mathbf{g}_k^t| \leq 6$. The extended-standard algorithm scores surprisingly well considering the maps of the previous section, a result which can again be attributed to the action of non-detections.

Fixing the fraction of the map to revisit a priori is a straightforward way to design subsequent sampling strategies and is apparently quite robust to errors in the marginal posteriors. That said, more efficient strategies are likely to require

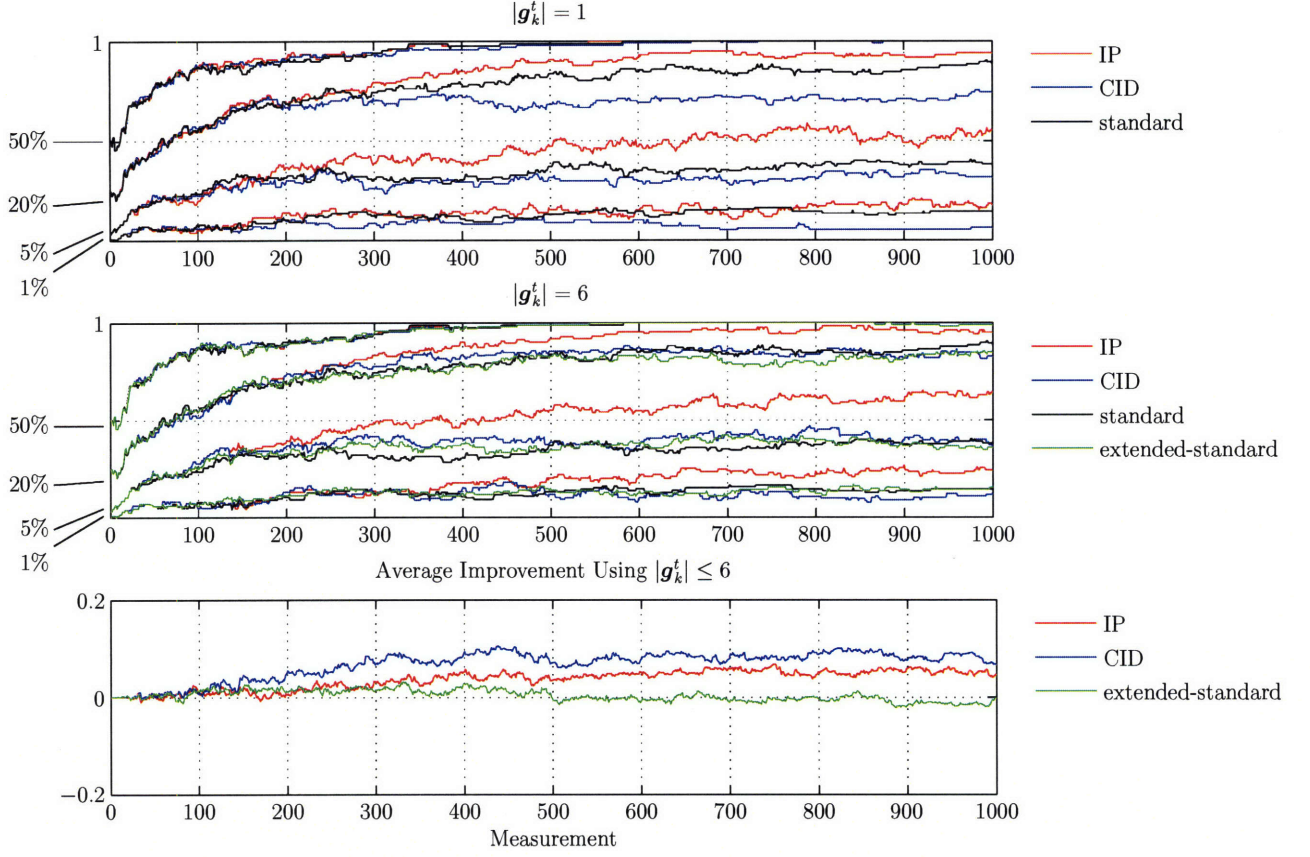


Figure 3-8: Percentage of the total number of occupied cells within the revisited map portion, averaged over 10 trials. The upper two panels show average scores versus time for each algorithm and various revisitation map fractions (indicated to the left of each panel). The bottom panel shows the average improvement attained using $|g_k^t| \leq 6$ across all revisitation-fractions for each algorithm.

accurate posteriors upon which to base real-time decisions. For instance, posteriors suggesting a high local density of occupied cells suggest a more valuable target than widely dispersed peaks because a sampling vehicle would need to spend less time transiting empty regions. Figure 3-7 supports the conjecture that the IP algorithm is capable of estimating posteriors consistent with the actual number of occupied cells. For the IP algorithm Fig. 3-9 investigates the behavior of the expected number of occupied cells within selected sub-regions of a map. The results show reasonably accurate and stable estimates of the actual number of occupied cells in each region after a few hundred measurements. With $|g_k^t| \leq 6$, these estimates fluctuate less. Unlike the maps produced by the other algorithms, these maps could be used to drive

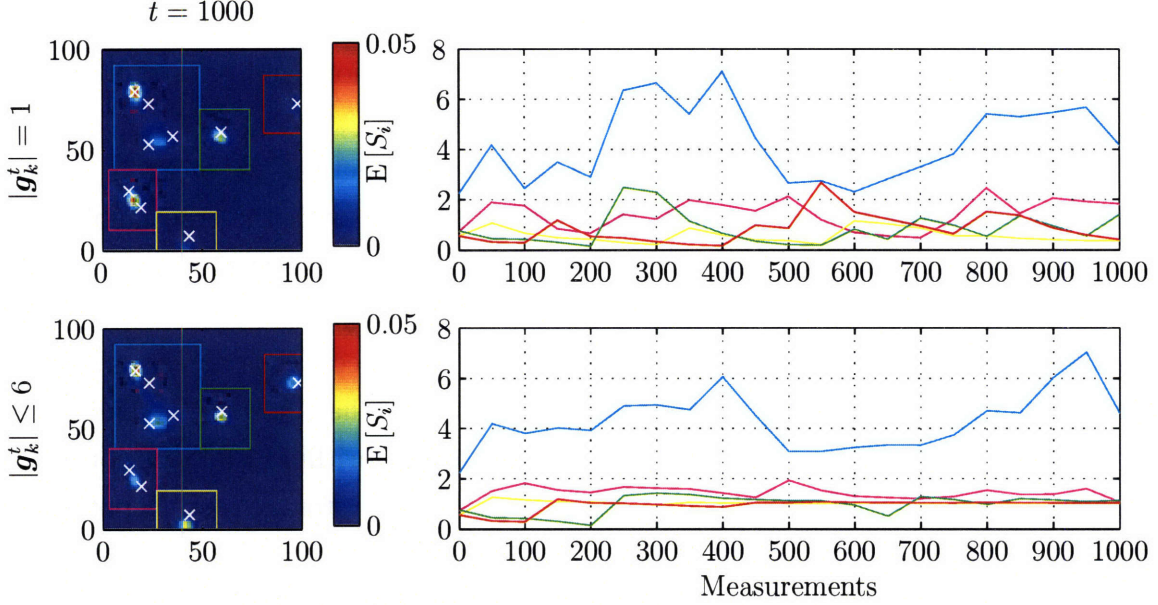


Figure 3-9: Expected numbers of occupied cells within selected regions of the map of Fig. 3-6 produced by the IP algorithm. The left panels show the OG maps after $t = 1000$ measurements for $\max |\mathbf{g}_k^t| \in \{1, 6\}$. To facilitate intercomparison, the colors indicate posterior probability of occupancy rather than $\log_{10} r_c^t$ as in the previous figures. The right panels show the expected number of occupied cells vs. time within the regions highlighted in the left panels. Both results are reasonably consistent with the actual number of occupied cells in each region suggesting the IP algorithm is capable of producing reasonable estimates of the true marginal posteriors. The results produced by processing subsets of detections ($|\mathbf{g}_k^t| \leq 6$) show reduced variance over those produced without using subsets of detections ($|\mathbf{g}_k^t| = 1$).

survey optimization strategies that trade-off transit time with the expected number of occupied cells within higher-resolution survey areas.

3.4 Summary, Contributions, and Future Work

I have devoted a large fraction of this chapter to exploring the appearance of inflated posteriors in OG maps produced in low prior environments as a consequence of the independence assumptions required. Indeed their appearance in some early simulations was the motivation behind the exploration of alternative OG algorithms that formed the subject of Part I of this thesis. Two classes of novel algorithms were introduced, based on: (1) partial relaxation of the standard CIM assumption required in standard OG mapping, or (2) the assumption that the marginal posteriors remain independent

even following detections. Figure 3-10 summaries the application domain of each algorithm studied in terms of the expected number of detections and prior probability of occupancy.

The IP algorithms hold the greatest potential for practical significance. They generated estimates of the posteriors that were consistent with the true number of occupied cells and produced better maps as the number of measurements increased. Furthermore, the improvements attained with IP by processing subsets of detections exactly with a maximum subset size $|\mathbf{g}_k^t| \leq 6$ instead of process each detection immediately upon reception ($|\mathbf{g}_k^t| = 1$) were modest suggesting that most of the enhanced performance of this algorithm relative to the standard algorithm in a low prior environment can be attained with no additional computation or storage cost. A potential weakness of this class of algorithm is the dependence of their results on the order with which measurements are acquired. Regular sampling trajectories may exacerbate this weakness and deserve further study. Nevertheless the successful applications of the IP algorithm to real world hydrothermal vent prospecting data appear in Ch. 5.

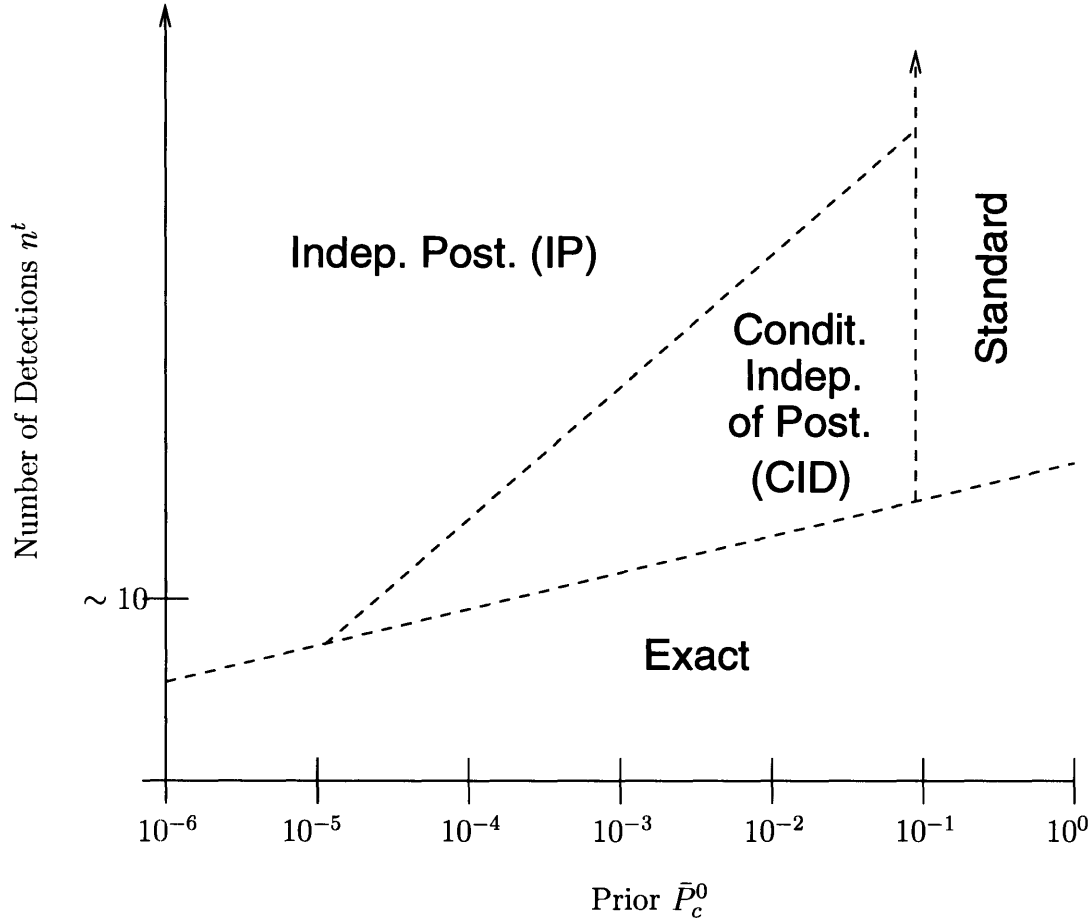


Figure 3-10: Binary measurement OG algorithm choice based on map prior and expected number of detections showing heuristic domains of applicability for each. The standard algorithm is appropriate for priors near 0.5. Other algorithms produce more accurate results for lower priors. For a small number of detections n^t an exact solution can be computed. Slightly larger numbers of detections can be handled by the CID algorithm. The IP algorithm performs well in low prior environments for moderate to large numbers of detections. Of course, these boundaries are subject to the parameters of the forward model employed. Higher probability of false alarm tends to blur the differences between the various algorithms because it reduces the correlation between detections, whereas increased correlation as from a wider sensor aperture (footprint) and tends to exacerbate the differences.

Part II

Automated Nested Survey for Hydrothermal Vent Localization and Mapping

Chapter 4

Hydrothermal Plume Survey by Autonomous Underwater Vehicle

Recent expeditions utilizing the Autonomous Benthic Explorer (ABE) AUV have demonstrated the effectiveness of an AUV for meso-scale vent localization [39, 65]. This chapter explores the hydrographic data attained by ABE during these expeditions and develops constraints on seafloor vent location from detections of hydrothermal effluent registered by the vehicle at various heights above the seafloor. AUV data is distinct from the more typical CTD-derived data usually employed for vent prospecting (e.g., [4]). Whereas towed CTD packages generate vertical slices through the plume, AUVs are more suited to producing densely sampled horizontal slices. Such surveys are more likely to intercept the relatively small signature of a buoyant plume (BP) emanating from a seafloor vent from within the much larger spatial signature of the so-called non-buoyant plume (NBP) that forms at some height above the seafloor as a result of persistent venting.

Depending on the tracer, hydrothermal plumes can be detected at a range of several kilometers in situ and out to thousands of kilometers with laboratory analysis of water samples [78]. Hydrothermal plumes inhabit a broad range of spatial and temporal scales, only a subset of which are resolvable by a surveying AUV, and not all of which contain useful information about the source location of that plume. A persistent theme throughout this chapter is to identify which scales are resolvable on the time and length scales imposed by an AUV survey, and subsequently to quantify the uncertainty inherent in the spatial relationship between water-column plume

effluent and seafloor source location.

The first contribution of this chapter is a methodology for the classification of certain hydrographic data collected by an AUV as indicative of background, NBP, or BP water. A necessary precondition for successful localization of hydrothermal vents is the identification of plume effluent in vehicle sensor data and the interpretation of these data in terms of the constituent components of a hydrothermal plume. Depending on a tracer's water column residence time and the predictability of its evolution, the presence of certain tracers can constrain the age of the plume being observed. I describe the processing applied to four tracers measured on ABE (potential temperature θ , acfOBS, reduction-oxidation potential (eH), acfVVA), each of which exhibits different characteristics depending on the stage of plume evolution.

The second contribution presented in this chapter is a model for the probable location of a buoyant plume's source on the seafloor parameterized by vehicle height at the time of detection. Background currents, tidally forced or otherwise are a significant environmental influence affecting hydrothermal plume structure. With the aid of an on-board acoustic Doppler current profiler (ADCP) and an accurate estimate of its own velocity-over-ground and heading, an AUV can measure water column velocities. The degree to which these measurements will provide constraints on source location depends on the degree to which the measured current record represents the temporal scales resolvable in the plume, the relative importance of background crossflows in defining plume structure, and the degree to which those influences can be modeled. The model enables an assessment of the refinement in the source location attainable by incorporating measurements of ADCP-derived crossflow velocity.

The development in this chapter considers only the information contained in single water-column detections in isolation, without regard to the spatial arrangement of multiple plume detections or regions in which no anomalies were observed. The occupancy grid (OG) mapping algorithms of Part I provide a methodology to generate maps of seafloor source locations consistent with all measurements; those results form the subject of Ch. 5. The current chapter develops the necessary background, specifically the probabilistic mapping between source location and water column detections, and methods to distill raw sensor data into the binary detection/non-detection form required by the methods of Part I.

This chapter is organized as follows: Section 4.1 provides a brief overview of the

physics governing hydrothermal plume evolution as they pertain to AUV-based survey. Section 4.2 describes the 3-stage nested survey methodology employed to localize hydrothermal vent sites with the ABE AUV. Section 4.3 presents the methods applied to raw hydrographic data from these surveys to identify hydrothermal anomalies and presents maps of hydrothermal plume activity. Section 4.4 demonstrates measurement of background current profiles from a vehicle-mounted acoustic Doppler current profiler (ADCP). Section 4.5 presents a conceptual model for short-timescale plume evolution, which I then employ to predict the utility of measuring current. The unknown parameters of this model are determined by comparison to successfully localized seafloor source locations.

4.1 Hydrothermal Plumes from an AUV’s Perspective

Hydrothermal vents occur where volcanic and tectonic processes induce the circulation of seawater through young oceanic crust.¹ Hot, chemically altered seawater is persistently discharged, or vented, as a turbulent, buoyant plume (BP) that rises above the sea floor to a height of 100–400 m in typical hydrographic settings [110], and then spreads laterally along isopycnals as the so-called non-buoyant plume (NBP). This structure suggests a natural strategy for “homing-in” on the source of venting: establish contact with the large spatial signature of the NBP; find the buoyant stems within, and finally follow these to the seafloor.

The following two sections provide a brief overview of the structure of and scales inhabited by hydrothermal plumes and outline the physical processes that drive their evolution.² I focus here on the implications for robotic searcher. Specifically, what

¹Not all hydrothermal venting is driven by a magmatic heat source. In particular, an exothermic reaction between seawater and young oceanic crust drives recently discovered hydrothermal venting at the Lost City site on the Mid-Atlantic Ridge [55].

²For the interested reader, McDuff [82] and references therein review physics of hydrothermal plume evolution in the buoyant phase through the use of classical models of buoyant plume evolution [89]. Lavelle [66] employs a large eddy numerical simulation to investigate the effects of crossflow. Middleton and Thomson [84] also investigate crossflow, but from an integral model perspective. The physical processes governing evolution of the NBP are more varied on account of the larger range of scales occupied by the NBP. Lupton [78] provides an overview; Helfrich and Speer [46] define the scales involved; Wetzler et al. [127] study the influence of ambient currents; and Thomson et.

physical processes affect the plume as mapped through the vehicle’s trajectory and visualized through the AUV’s sensors?

4.1.1 Buoyant Hydrothermal Plumes

The dynamics of typical buoyant hydrothermal plumes are dominated by turbulent entrainment [119] of the surrounding seawater. This process dilutes vented hydrothermal effluent by a factor of 10^4 – 10^5 by the time the density of the mixed fluid matches that of the surrounding fluid at its terminal rise height [77]. Turbulent entrainment of ambient waters makes regional hydrography an important influence on the trajectory followed by an ascending BP, and acts directly to obscure source location by expanding plume diameter during ascent. Ambient crossflows also influence plume trajectory by bending the plume centerline in the direction of flow.

Turbulent Entrainment In a time-averaged sense, over the course of its rise from the seafloor a buoyant hydrothermal plume will have expanded laterally from on the order of a few centimeters at an individual vent orifice to a diameter of 50 – 100 m at equilibrium height based on model predictions [110]. Instantaneously, the buoyant plume is resolvable as a patchy, irregular structure of high intensity anomalies interspersed with unaltered background water [78]. This irregular structure is the result of the turbulent entrainment of surrounding water as the plume ascends: the buoyancy of the discharged vent water creates a strong vertical instability which in turn sets up shear instabilities at the edges of the ascending fluid resulting in eddies that engulf ambient fluid and mix it into the interior of the plume [119]. From the perspective of a robotic searcher, this implies time-averaged quantities like plume width and mean tracer concentration, which carry information about distance-to-source and source intensity and which could define a concentration gradient toward the source, are difficult to resolve. However, detection is facilitated by both the much higher than average tracer intensities encountered, and by the rapidly fluctuating signal, characteristics which become more pronounced as the plume’s spatial signature decreases toward the seafloor (Fig. 4-1).

al [114] explore venting-induced circulation and the role of bathymetry.

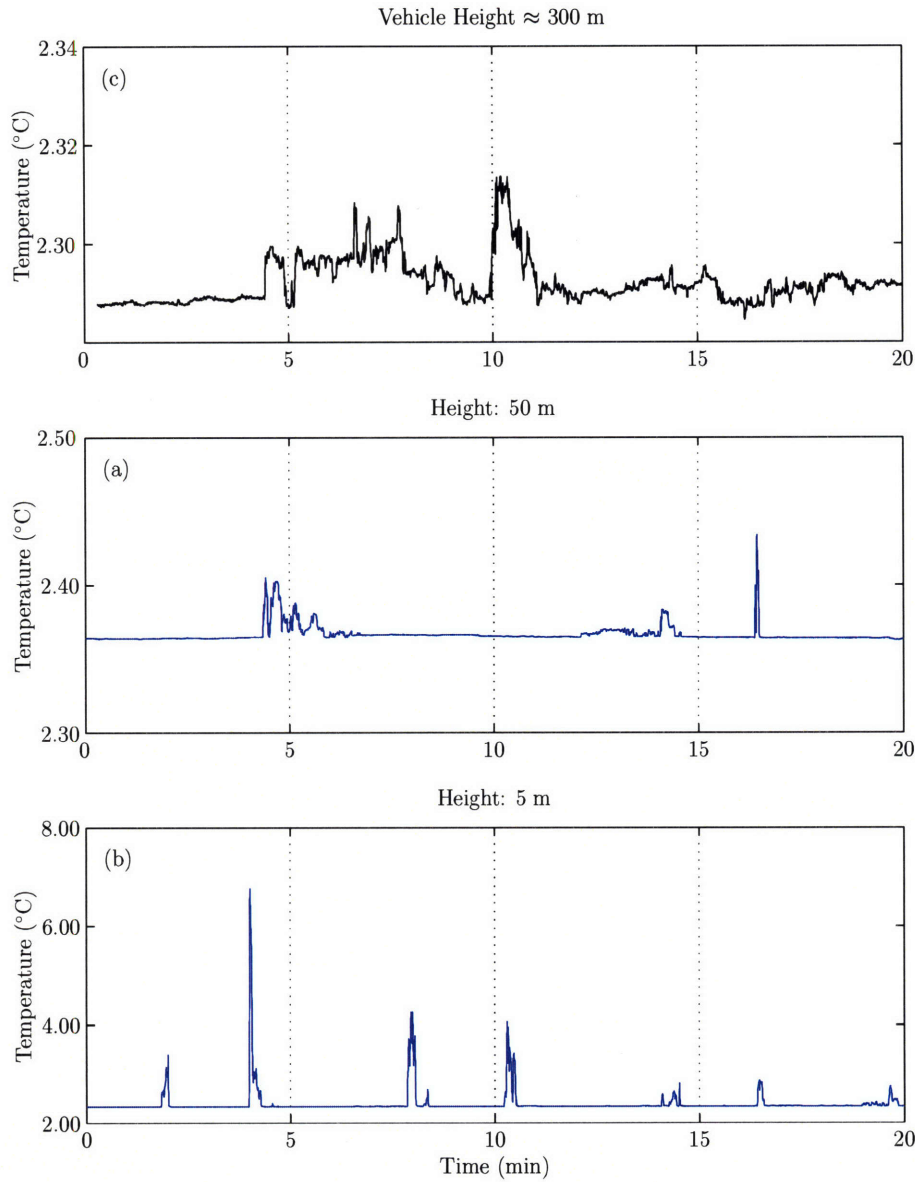


Figure 4-1: In situ temperature records collected by ABE as it passed at approximately constant speed (0.6 m/s) through buoyant hydrothermal plumes emanating from the Kilo Moana vent field on the ELSC (20°3' S, 176°12' W): **(a)** 300 m above bottom at the depth of the NBP; **(b)** 50 m above bottom; **(c)** 5 m above bottom. Plots (b) and (c) indicate repeated contact with individual nearby plume stems, whereas plot (c) shows a relatively broad peak that may be the result of plumes from individual sources below having coalesced. In all cases the turbulent nature of buoyant plumes is evident in the fluctuations of the signals, but its spatial extent increases with height off-bottom while the magnitude of the observed temperature anomaly decreases by two orders of magnitude. Time series like these are characteristic of vent-prospecting surveys conducted by ABE. Additional examples are provided in App. F.

Regional Hydrography Regional hydrography, particularly the strength of the (stable) density gradient, plays a critical role in determining terminal rise height. Steeper gradients result in lower rise heights because the density deficit of the rising plume relative to background water at the same height decreases more rapidly with height due to both the steepness of the background density gradient and the entrainment of relatively denser waters during ascent. In contrast, the dependence of rise height on source heat flux over the range of typical black smoker vents (1 MW to 100 MW) is relatively weak, producing less than an order of magnitude difference in terminal rise height for the same density gradient [78]. Temperature and salinity profiles (not just the resulting density gradient) are also important. Though both the deep Atlantic and deep Pacific are stably stratified, the salinity gradient in the Atlantic decreases with depth, whereas it increases in the Pacific. Curiously, this difference produces a warm and salty NBP relative to background in the Pacific and a cold and fresh NBP in the Atlantic [110]. Furthermore, source chemistry (salinity) affects both rise height and the temperature anomaly associated with terminal rise height [82], to the extent that expected sign of the anomaly changes (e.g. [108]). The dominant role of regional hydrography is fortuitous from the perspective of a robotic searcher because background profiles can be determined a priori (from CTD casts) or measured by the vehicle before commencing a search (e.g. during its descent).

Ambient Crossflow Background currents impart their momentum on the rising water in buoyant plume stems, thereby bending the plume's centerline in the direction of the crossflow. Acceleration to crossflow velocity is complete within the immediate vicinity of the source [84]. Crossflows result from the combined influence of tidal currents, basin-scale circulation, episodic events, and circulation induced by the venting itself [114]. Crossflow velocities at hydrothermal vent sites vary: Data acquired by Thomson et al. [113] on the Endeavour Segment of the Juan de Fuca Ridge reveal tidally-forced oscillatory currents of up to 10 cm/s with superimposed steady currents of 5 cm/s and strong attenuation of these currents within the confines of the axial valley. The maximum crossflow magnitude observed by ABE from on-board ADCP measurements as of 2006 was 16 cm/s at a site on the Southern Mid-Atlantic Ridge (SMAR) (Fig. F-73). The location of buoyant plumes high in the water column reflects the combined effect of turbulent entrainment and superimposed

advection. For a characteristic rise time of 1 h [110], a crossflow of 10 cm/s will advect some effluent a few hundred meters from its source. Crossflows also reduce rise height by as much as a factor of two for very strong crossflow magnitudes of 20 cm/s. [84].³

Other Factors Other factors affecting the behavior of hydrothermal plumes in the buoyant phase are source geometry, entrainment of nearby diffuse venting, alteration of the background profile by persistent venting, and density alteration on account of the presence of suspended particles [82]. Nearby plumes from multiple sources may also coalesce as their expanding cross-sections interfere with one another.

4.1.2 Non-buoyant hydrothermal plumes

The laterally-spreading non-buoyant plume evolves according to a self-imposed pressure gradient, the influence of background currents (mean, tidally forced, and venting-induced), and beginning at scales $O(1000 \text{ km})$, by the rotation of the earth [46]. The vertical thickness of the non-buoyant plume $O(100 \text{ m})$ is dependent upon variations in the intensity of background currents which alter the rise height of buoyant plumes by inducing lateral drift during ascent and thereby increasing the amount of mixing achieved for a given rise height.

At the point of neutral buoyancy, the remaining vertical momentum of the rising plume water dictates that it will overshoot before sinking back into the non-buoyant layer. This fact has important implications for AUV-based search because it implies buoyant stems completely penetrate the non-buoyant plume, and can be detected by an AUV surveying within its vertical confines.⁴

Once the depth of the non-buoyant plume has been established, an AUV offers certain advantages over a towed package as a means of finding the relatively small scale $O(100 \text{ m})$ signature of buoyant stems within the non-buoyant plume. AUVs are highly maneuverable, relatively fast, and can navigate with a precision on the

³Rudnicki and German [108] report field evidence of rise-height modulation by crossflow speed from a continuous record of a stationary vertically-oscillating CTD cast.

⁴This simplistic perspective is strictly true only for a single BP in an otherwise steady environment. Multiple sources and dynamic crossflows can create non-buoyant layers over a range of a few hundred meters depth, sufficiently thick that not all buoyant stems will necessarily penetrate this layer completely.

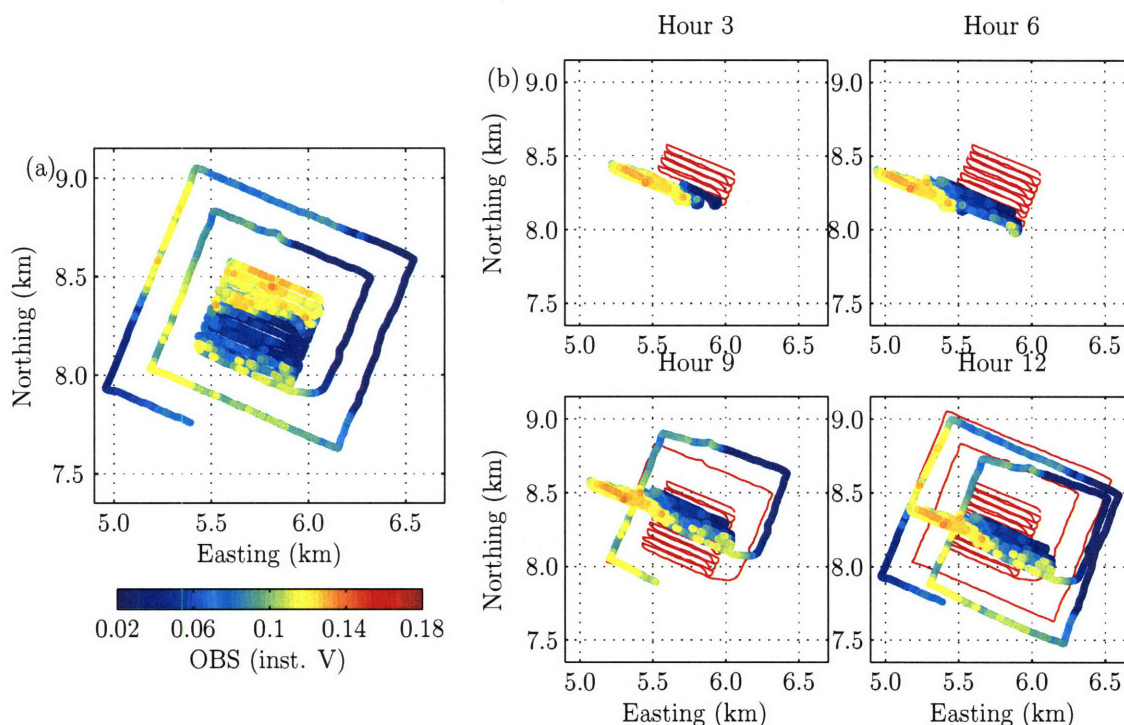


Figure 4-2: Optical backscatter (OBS) data collected by ABE over High Rise vent field on the Juan de Fuca Ridge ($47^{\circ}54' N$, $129^{\circ}10' W$). The left plot shows OBS plotted according to the position of the vehicle at the time each sample was taken, i.e. as if the data were acquired synoptically. In reality, the vehicle commenced surveying in the upper right corner of the inner high-density grid and completed its survey on the outer arm of the spiral 12 h later. The image is not representative of the true state of the plume at the conclusion of the survey: the sequence of images on the right show snapshots of survey data during the dive with the position of each OBS measurement adjusted for integrated current velocity from the time the sample was taken. The vehicle's track is shown in red. During the first part of the survey, ABE re-sampled essentially the same plume water repeatedly as it advected to the southwest. Credence to this interpretation of data is provided by the consistency between OBS measured on tracklines passing through the predicted positions of old measurements.

order of meters with the addition of external acoustic beacons [134]. The usual trade-offs between coverage, resolution and synopticity [129] still apply of course, but for survey areas $O(\text{km}^2)$, AUVs can produce plume maps that are spatially of higher resolution (precise tracklines)⁵, and perhaps also temporally closer to synoptic (higher speed). Additionally, water velocity profiles from ADCPs, which are becoming standard instrumentation on AUVs, can be used to compensate maps for asynopticity, as was done to create the sequence of OBS maps in Fig. 4-2. Finally, and perhaps most significantly, AUVs have sufficient computational resources aboard to enable data-driven adaptive surveys [12, 31, 134].

Buoyant plume stems are discernible within the non-buoyant plume both physically and chemically. Sufficiently close to the centerline of a rising plume stem, the net upward transport of water is often sufficiently strong to overcome ABE's vertical thrusters resulting in a forced excursion from desired depth.⁶

Currents with strong tidal components complicate the process of inferring the locations of buoyant plume stems directly from measurements of essentially conserved tracers (e.g. potential temperature). Small mean components dominated by tidally forced oscillations create a plume structure that is simultaneously variable on tidal timescales and yet represents the integrated output of several tidal cycles worth of discharge (see particularly [120, 127]). As a result, locally high concentrations, especially of conservative tracers like potential temperature, resulting from periods of slack tide can occur far from buoyant stems. Records from multiple tracers subject to differing reaction rates with seawater can constrain the age of non-buoyant plume water [53]

⁵Both towed assets and AUVs can be acoustically navigated in an absolute frame to yield an accurate map; however, because they lack the maneuverability that is available to AUVs, plume maps produced with towed assets are limited in resolution.

⁶Previous expeditions to hydrothermal sites on the Juan de Fuca Ridge with ABE have employed an acoustic travel-time velocimeter to directly measure vertical velocity in support of hydrothermal heat-flux measurements [111, 132].

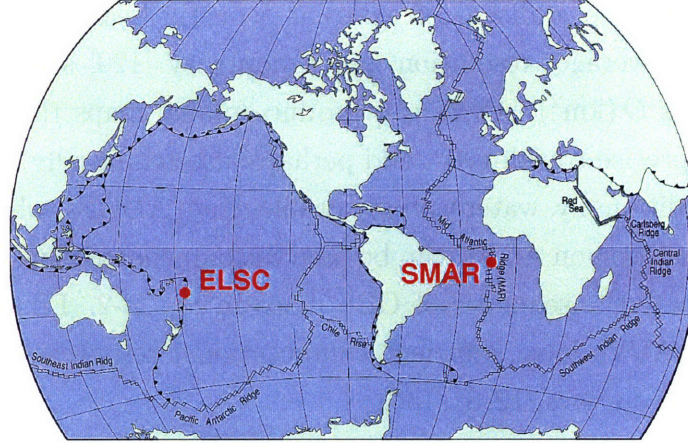


Figure 4-3: The mid-ocean ridge system, and the locations of two recent hydrothermal vent-finding expeditions that utilized an AUV as complementary to towed assets. ELSC ($20^{\circ}3' S$, $176^{\circ}12' W$): Eastern Lau Spreading Center (September/October 2004). SMAR ($4^{\circ}54' S$, $12^{\circ}28' W$): Southern Mid-Atlantic Ridge (March 2005).

4.2 Background: A Three-Stage Nested Survey Approach

On two separate expeditions, to the Eastern Lau Spreading Center (ELSC) (September/October 2004) and to the Southern Mid-Atlantic Ridge (SMAR) (March 2005) the AUV ABE [133] was employed to localize the undiscovered sources of several known hydrothermal plumes and to provide preliminary characterizations of the associated vent fields upon discovery (Fig. 4-3). Critically, in both cases ridge-scale hydrographic and bathymetric work carried out using towed and lowered assets had previously identified promising sites (ELSC: [80]; SMAR: [39]). This type of data is essential for initially establishing contact with the NBP, as the task of finding $O(10 \text{ km})$ plumes on the ridge-scale $O(100\text{--}1000 \text{ km})$ remains beyond the energy capacities of most AUVs.

The AUV performed three successively finer-scale nested survey dives at each site, according to the methodology first deployed in the ELSC [39, 65]:

- Phase-1: Hydrographic mapping of the NBP (200–400 m altitude)
- Phase-2: Hydrographic and multibeam bathymetric mapping (50 m altitude).
- Phase-3: Photo-mosaicking of individual vent fields (3–5 m altitude).

Each stage was executed at progressively lower altitudes and finer trackline spacing based upon the results of the previous stage and any available ancillary data, a process that broadly reflects the physical structure of hydrothermal plumes themselves (§ 4.1). Ideally, Phase-1 hydrographic data was sufficient to locate BP stems intersecting the NBP, or at a minimum to better constrain regions of highest intensity anomalies.⁷ Assuming conclusive results from Phase-1, Phase-2 hydrographic data usually contains multiple BP interceptions, further constraining the extent of any vent fields. Phase-3 data then serves to map individual hydrothermally active regions within a vent field on a scale of meters and via high-resolution digital still photography to provide preliminary data for biological characterization of the site. Figure 4-4 shows an example of tracklines from Phase-1 through Phase-3 at an ELSC site.

Most of the data in this chapter was acquired over the course of several Phase-1 through Phase-3 nested survey progressions. The dataset is unique in that all of it was collected by a single vehicle (ABE) and includes both water-column plume data (from Phase-1 and Phase-2 dives) as well as groundtruth vent locations (from Phase-3 dives). Given the apparent success [39,65] of the strategy outlined above, one might legitimately question the value of further analyzing the data to improve search strategies. After all, at least some vent fields were located at most sites following the discovery of NBP activity by other means.⁸ On the contrary, there is a great deal to be gained from careful a posteriori analysis of this data, particularly with an eye toward automation of the above procedure on future expeditions:

reliability: On occasions where no vent fields were discovered on Phase-3 dives, could automatic interpretation of data have indicated that Phase-1 and/or Phase-2 data was insufficiently conclusive to warrant progressing to the next stage?

efficiency: How efficient (in terms of trackline length) was human design of Phase-2 and Phase-3 dives, i.e., was the uncertainty in source location based on water

⁷It was already remarked upon in § 4.1.2 that O (km) structure is strongly influenced by oscillatory tidal currents which can produce strong anomalies not directly associated with the instantaneous locations of buoyant feeder plumes.

⁸ELSC: successful localizations at 3 out of 5 sites; SMAR: successful localization of two distinct vent fields at one site.

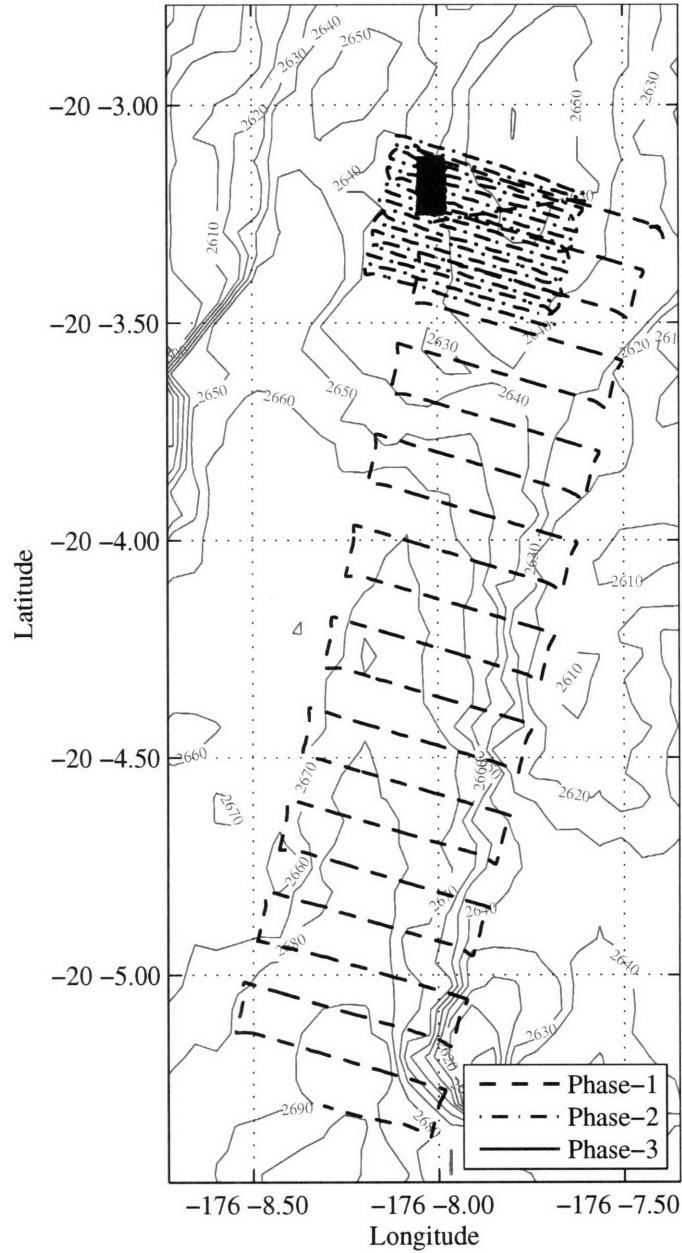


Figure 4-4: Vehicle tracklines from three nested surveys (Phase-1, Phase-2, Phase-3) at a hydrothermal site on the ELSC ((20°3' S, 176°12' W)).

column detections over or underestimated?

autonomy: Most importantly, could this strategy be carried out autonomously by allowing the vehicle to design successive phases of the dive based on its own

autonomous interpretations of the data so as to make better use of available bottom time?

To attain answers to these questions, interpretation of the raw data from Phase-1 and Phase-2 has to be made automatic and quantitative. That is the goal of the next three sections. Lest the potential savings associated with automation be underestimated, consider that the ratio of time associated with inter-survey recovery, human analysis of the data acquired, vehicle maintenance, and re-deployment relative to the time spent actually collecting data on these expeditions was roughly 2-to-1.

4.3 Anomaly Maps

A necessary component of any strategy for finding hydrothermal vents based on the locations of water-column plume detections is reliable classification of water-column data into background water unaffected by hydrothermal input and plume water. Hydrographic sensors measure absolute concentrations of chemical or physical quantities, not the presence or absence of hydrothermal effluent. Background values must be subtracted from raw measurements to arrive at the portion of the signal that represents hydrothermal input. This would be trivial if background concentrations were known perfectly. In reality, many tracers have background concentrations that are subject to variability in space and time, and inevitably suffer some contamination from sensor noise.

In this section I develop a methodology for the detection of hydrothermal effluent from sensor measurements collected by an AUV. I demonstrate its broad applicability by presenting results using four markedly different tracers recorded by the ABE AUV (described subsequently) during several recent vent-prospecting missions.⁹ I do not claim the methodology presented here is optimal, just that it is both effective and practical, particularly in that the number of tunable parameters is deliberately kept small. Different strategies are required for detection of effluent in the non-buoyant plume (NBP) versus buoyant plume (BP) stems on account of the vastly different spatial scales inhabited by each. Fortunately, this separation in scales also enables

⁹Complete results for ABE dives 128, 131, 136, and 137 (ELSC); and dives 150, 151, and 152 (SMAR) appear in App. F. In this section, I restrict my attention to ABE-126, a Phase-1 dive at ELSC that included interception of a buoyant plume.

reliable detection of BP stems from within the NBP, the key component of the three-phase survey strategy outlined in § 4.2.

NBP The approach to detection of the NBP is essentially analogous to the standard techniques applied to lowered or towed CTD data (e.g. [4, 77, 78]), with two enhancements tailored to autonomous operations:

1. Self-contained: background tracer profiles are determined using data collected during vehicle descent.
2. Classified output: continuous-valued anomalies are classified into statistically significant detections and non-detections of plume water based on deviation from the profiles determined during vehicle descent.

The first of these eliminates the need for cross-calibration between sensors aboard the AUV and other platforms and minimizes the impact of spatial and temporal variability in background tracer concentration. The second allows delineation of the plume and could potentially provide the necessary input to a stochastic mapping algorithm like that applied to BP detections subsequently in Ch. 5.¹⁰ As the goal is to disambiguate NBP water from unaffected background water, the method applies only to conservative tracers or non-conservative tracers (see § 4.3.1 for a definition of these terms) with long residence times. That said, non-conservative tracers with relatively short residence times have proven extremely useful in qualitatively constraining the location of BP stems from Phase-1 ABE data in the event of no definitive interception of buoyant effluent. I return to this point in the conclusion of the present chapter.

BP In contrast to a typical NBP resolvable in situ at kilometer scales, typical buoyant plume stems represent a target orders of magnitude smaller in horizontal extent. Whereas a Phase-1 survey conducted with ABE may remain in contact with a NBP over most of the dive, BP detections are rare. Consequently methods for BP declaration based on outlier detection are suitable, whereas they would likely fail if applied to NBP detection. An outlier-based approach avoids the need for an assumed distribution to describe measurements while in contact with BP effluent.

¹⁰The potential payoff is significant as the scales over which effluent in the NBP is detectable in situ far exceeds the spatial scale of BP stems within the NBP.

That feature is attractive, as the turbulent structure of a BP implies a strongly fluctuating signal whose distribution would be difficult to predict in any case, and especially so without a priori knowledge of source parameters. The vigorous and relatively unpredictable fluctuations in tracer concentration within a BP strongly favor an outlier-based approach to detection.

Ultimately the outlier-based approach still comes down to the careful choice of a threshold. The principal challenge to such an approach is variation in the background (potentially significant in hydrothermally active areas) which does not necessarily obscure BP activity so much as make choosing a threshold a potentially brittle undertaking. Automatic threshold determination is desirable and should minimize the number of parameters requiring manual tuning. The approach pursued herein first removes predictable trends in tracer concentration before batch-processing each tracer to automatically assign a meaningful threshold. The last steps classify measurements into binary detections and non-detections according to their status as outliers, and finally combine these binary classifications into a single binary indicator variable based on a heuristic assessment of the specificity of each tracer. Batch-processing trades the potential for real-time detection (and perhaps triggering of special vehicle actions) for increased robustness, in line with the philosophy of a nested survey approach to search.

4.3.1 A Taxonomy of Hydrothermal Tracers

Hydrothermal effluent contains a wide variety of chemical species that react at varying rates with seawater so that the chemical signature of a plume is dependent in part on water-column residence time, but also on source water composition, background water composition, and biological factors [73]. At the coarsest level, hydrothermal chemical hydrographic tracers are of two types [78]:

conservative tracers whose concentrations are affected only by passive advection and diffusion, and

nonconservative tracers whose concentrations are affected in addition by chemical reactions, biological processes, or radioactive decay.

The most readily measured conservative hydrothermal tracers are the physical properties temperature and salinity. Though the density of a NBP is equivalent to

the background value by definition, the potential temperature and salinity within this layer deviate from background profiles [77]. As mentioned previously, the sign of these anomalies is dependent on background stratification, resulting in a warm and salty hydrothermal anomaly in the Pacific and, counter-intuitively, a cold and fresh hydrothermal anomaly in Atlantic waters [110].

Of the non-conservative chemical species enriched in hydrothermal source fluids, hydrogen H_2 , methane CH_4 , iron Fe, and manganese Mn have received the most attention [5, 73], and many can now be measured in situ on packages suitable for AUV deployment [13, 17, 96, 105, 112]. While some of these new instruments have seen developmental deployment on ABE [38], that data will not be considered here. Instead, I focus on data from two instruments that provide proxy measurements related to the presence of these chemical species: optical backscatter (OBS) and reduction-oxidation (redox) potential (eH). OBS is related to some of the non-conservative chemical tracers listed above by virtue of the particulate matter they form as a result of chemical transformations initiated upon contact with seawater. Redox potential provides an indication of the age of nascent plume water by measuring the degree to which these and other reactions have progressed.

The chemical evolution of hydrothermal particulate matter depends on oxidation, precipitation, dissolution, adsorption and scavenging reactions, many of which are biologically mediated [53, 73]. Particulates settle out from the water column at varying rates to form distinctive deposits near hydrothermal sites, or slowly dissolve back into the water column once mixing has led to sufficient dilution with ambient seawater. Optical backscatter and transmissivity provide sensitive and economical optical proxies for particle concentration and are increasingly used to infer the presence of hydrothermal venting without other indicators [5, 7]. The complexity of the physical and chemical transformations undergone by hydrothermal particulates as well as the particle-size-dependent nature of optical measurements makes these measurements difficult to interpret in terms of source proximity. Nevertheless, with the exceptions of false positives due to sediment resuspension and missed detections of diffuse or low temperature venting, optical properties of the water column “almost invariably provide a reliable indicator of underlying hydrothermal activity” [5].

The reduction-oxidation (redox) potential (eH) of an aqueous solution is a measure of the affinity of dissolved chemical species for acquiring electrons, a process known as

reduction. Measured in volts, eH is expressed relative to the redox potential of hydrogen, which is set to zero by definition. Substances with positive redox potentials are more capable of oxidation than hydrogen; substances with negative redox potentials are more capable of reducing. Hydrothermal fluids tend to be enriched with reducing chemical species from circulation through young oceanic crust prior to emission from vents as plumes. Contact with seawater will oxidize these chemical species, though at varying rates [53]. Thus anomalously low eH is indicative of recently expelled hydrothermal effluent [95].

4.3.2 Anomaly Definition

I use the term *anomaly* to refer to the signed difference between the value of some scalar quantity v and its background value v_0 measured at identical reference conditions \mathbf{y} . In notation of [120], $\Delta_{\mathbf{y}}v = v(\mathbf{y}) - v_0(\mathbf{y})$. (I reserve the label *detection* for anomalies that exceed some to-be-determined threshold.) Background values could be tabulated over the relevant ranges of the reference variables, or assumed to satisfy some functional relationship $v_0 = f(\mathbf{y})$. A complete description of the background would be given by a \mathbf{y} composed of time and the three spatial coordinates, though other choices of reference variables might prove more suitable. Indeed, most hydrographic parameters in the deep ocean vary strongly with depth and only weakly otherwise. An appropriate choice of reference variable(s) should reflect the expected complexity of the background relationship and also the fidelity with which that relationship can be known a priori or determined from measurements. Tracers specific to hydrothermal activity, and therefore lacking any dependence on reference variables, can be accommodated with the notation Δv .

Algorithm 8 outlines the entire procedure applied to raw sensor measurements. These are assumed to fall into two classes: (1) tracers exhibiting variation with depth, and (2) tracers specific to hydrothermal venting. Raw data are first pre-processed in a way specific to each sensor (detailed in the text for the tracers studied herein). The next steps remove trends due to background profiles, and in the case of BP-detection, also contamination from background profile alteration (for surveys below the NBP) and the NBP (for surveys within the NBP). The final steps generate a binary output on the basis of an automatic threshold determination. These steps are

different depending on whether detections should reflect NBP or BP contact. In either case, final declarations of NBP contact and BP interceptions are made by combining indicator variables from k tracers according boolean functions $f : \mathbb{B}^k \rightarrow \mathbb{B}$.

The key computations involved in each step are provided as pseudo-code along with a flowchart to indicate signal flow. Most of the processing applied to NBP detection applies to BP detection as well. The additional aspects of BP detection are designed to remove low-frequency trends prior to outlier declaration. Diagonal open-headed arrows in the flowchart indicate key user-supplied information and tunable parameters. These amount to a single scalar P_F to denote a desired probability of false alarm (possibly different for each tracer), and the boolean functions f_{NBP} and f_{BP} . The following sections discuss each step in detail, beginning with the preprocessing applied and specific to the tracers measured on ABE. The remaining steps of the flowchart should be readily applicable to other measurements beyond the ones studied here.

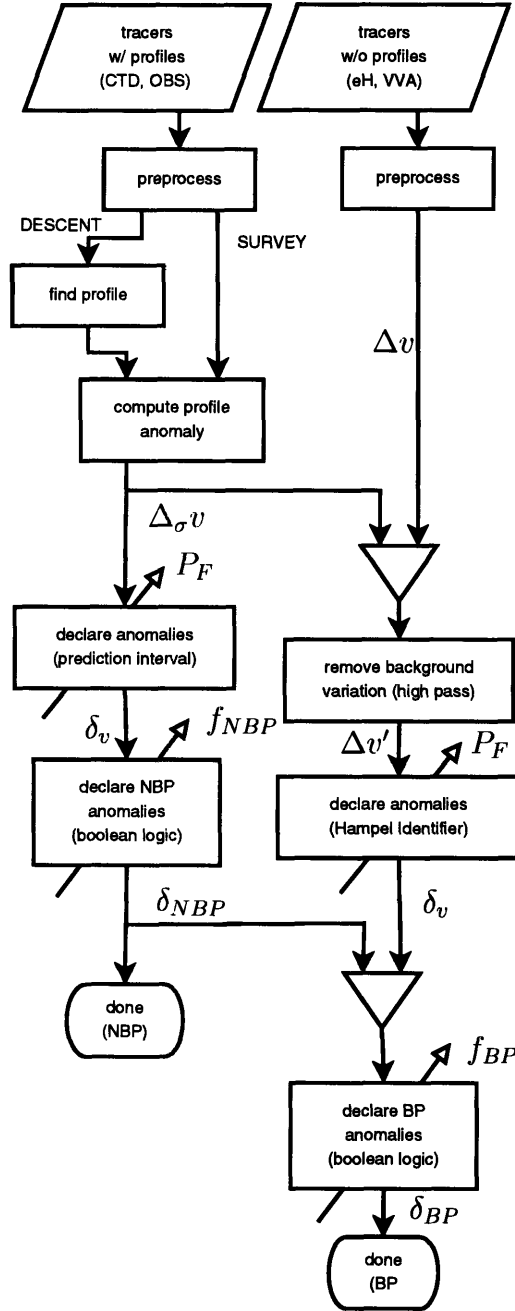
4.3.3 Preprocessing Applied

The paragraphs that follow detail the specific processing applied to a representative subset of sensors aboard ABE used for hydrothermal plume detection: potential temperature θ , optical backscatter (OBS), reduction-oxidation potential (eH), and vertical velocity anomaly (VVA). This selection spans conservative to highly nonconservative behavior, and includes tracers both with and without background profiles. The specific preprocessing applied in each case is unique but may include any/all of the following: combination of raw sensor measurements into a composite measurement, removal of certain sensor pathologies, and pre-filtering. All measurements discussed below were sampled between 1 Hz and 3 Hz.

Potential Density In this study, potential density σ is not used as a tracer directly,¹¹ instead serving as a reference variable for tracers with profiles; however it too is a measurement subject to error. Potential density is computed from in situ temper-

¹¹The vertical gradient in potential density, measured for instance by two vertically-displaced conductivity/temperature probe pairs, could be used as a hydrothermal tracer. The normalized gradient is known as stability [57]. Viers et al. [121] inferred negative stabilities (density inversions) to be indicative of BP contact.

Algorithm 8 Hydrothermal plume anomaly processing and detection algorithm and key function definitions. User-supplied inputs, indicated by open-headed arrows, are limited to specifying a probability of false alarm P_F and the function f . Relational operations in square brackets $[\cdot]$ produce a binary result.



Variable Definitions

v	tracer measurement
σ	potential density
$\Delta_{\sigma}v$	v -anomaly w.r.t. profile
Δv	v -anomaly
$\Delta v'$	v -anomaly following filtering
δ_v	binary indicator (v)
δ	binary indicator (BP)
P_F	tunable false alarm rate

Function Definitions

find_profile(v, σ) {

$$\frac{\partial \hat{v}}{\partial \sigma}, \hat{v}_{ref} = \underset{\frac{\partial \hat{v}}{\partial \sigma}, v_{ref}}{\operatorname{argmin}} \sum_t (v^t - \frac{\partial \hat{v}}{\partial \sigma} \sigma - v_{ref})^2$$

return $\frac{\partial \hat{v}}{\partial \sigma}, \hat{v}_{ref}$

}

compute_profile_anomaly($v, \frac{\partial \hat{v}}{\partial \sigma}, \hat{v}_{ref}$)

{

$$\Delta_{\sigma}v = v - (\frac{\partial \hat{v}}{\partial \sigma} \sigma + v_{ref})$$

return $\Delta_{\sigma}v$

}

prediction_interval($\Delta_{\sigma}v$)

{

$$\hat{\lambda}_{\Delta_{\sigma}v} = \text{sample variance } \Delta_{\sigma}v$$

$$\delta_v = \left[\frac{|\Delta_{\sigma}v|}{\hat{\lambda}_{\Delta_{\sigma}v}} > \Phi(1 - P_F/2) \right]$$

return δ_v

}

Hampel_Identifier($\Delta v'$) {

$$\Delta v^{\dagger} = \text{median } \Delta v'$$

$$S = 1.4826 \text{ median } |\Delta v' - \Delta v^{\dagger}|$$

$$\delta_v = \left[\frac{|\Delta v' - \Delta v^{\dagger}|}{S} > \Phi(1 - P_F/2) \right]$$

return δ_v

}

ature, salinity and pressure using an empirical equation of state for seawater [36,88]. Temperature and pressure can be measured directly, but salinity is usually derived from measurements of conductivity. As with most lowered CTDs, the conductivity and temperature probes on ABE are very nearly co-located and a constant flow rate pump ensures that the sensors are flushed nearly simultaneously with the same parcel of seawater. Nevertheless, because salinity varies over a relatively small range, the conductivity of seawater tends to track temperature closely and even slight mismatch between the temperature signal and the portion of the conductivity signal due to temperature can lead to so-called “salinity-spiking” [97].¹² This is a cumulative effect due to temporal delay between the sensors; thermal mass of the conductivity cell; and a difference in diffusion rates between heat and salt [76]. Since we employ potential density as an independent variable, noise from the raw sensor measurements as well as from salinity-spiking contribute to a noise floor that is present in the determination of all subsequent anomalies.

Potential Temperature As with potential density, potential temperature is derived from an empirical relationship that depends on in situ temperature, salinity, and pressure [36,88]. Heat is conserved, therefore potential temperature behaves as a conservative tracer. As discussed in § 4.1, the sign of the potential temperature anomaly depends on background stratification and height above the source. In the Pacific, the $\Delta_\sigma\theta$ associated with hydrothermal venting is always positive. Non-buoyant hydrothermal plumes in the Atlantic result in $\Delta_\sigma\theta < 0$; however, at some distance beneath this level of neutral buoyancy, the sign of $\Delta_\sigma\theta$ associated with the buoyant plume becomes positive.¹³ On ABE, 99.9% prediction intervals for background $\Delta_\sigma\theta$ are typically 2.5 m°C. The maximum recorded in situ temperature to date was $\sim 60^\circ\text{C}$, observed during a near-bottom dive on the ELSC.

¹²Salinity spiking tends to plague vertical CTD casts more so than typical horizontal AUV surveys because vertical gradients of salinity and temperature in the ocean dramatically exceed horizontal gradients. However, strong horizontal gradients can be encountered in waters contaminated by hydrothermal venting.

¹³Modeling by Speer and Rona [110] for the time-averaged behavior of a typical Atlantic plume predicts a change of sign in $\Delta_\sigma\theta$ at a rise height of 200 m however, significant negative anomalies were observed on ABE dives 151 and 153 at only 50 m height.

Optical Backscatter OBS is measured on ABE with a sensor that outputs a signal proportional to the intensity of light backscattered from a small volume illuminated by the sensor.¹⁴ Spikes in the output are frequent, presumably the result of some larger-than-average particle momentarily occupying the sensing volume. To attenuate these high-frequency phenomena, the raw output is passed through a 10-point median filter. Few sources of suspended matter exist in the deep sea besides hydrothermal venting and consequently any dependence of OBS on depth is usually slight. This fact along with the relatively low cost of OBS instrumentation has made OBS a popular tool in hydrothermal prospecting, especially for identifying NBP waters. However, the complex dependence of particle formation on vent and background water chemistry and particle fallout renders quantitative interpretation of OBS difficult.¹⁵ Only positive deviations from background are considered potentially indicative of hydrothermal venting. OBS measurements on ABE following median filtering typically fall within 2 mV of the mean (instrument voltage, 5 V full scale). The OBS data presented subsequently have been corrected for a consistent 0.1 mV/h drift, presumably the result of a slowly falling supply voltage to the instrument's internal voltage regulator as ABE depletes its batteries during a mission.

Redox Potential On recent expeditions, ABE has carried a reduction-oxidation potential (eH) probe provided by Dr. Ko-ichi Nakamura[†] that has enabled this valuable measurement to be made in situ. The sensor has a complicated response characterized by a relatively fast onset time and slow recovery time of hundreds of seconds. Ongoing work by the instrument's inventor suggests that the magnitude of differentiated eH ($d/dt(\text{eH})$) is correlated with the concentrations of reducing chemical species typically encountered in vent fluids [94]. Based on experience, steeply decreasing eH

¹⁴Seapoint Sensors, Inc., <http://www.seapoint.com/>

¹⁵For instance, iron-oxides make up a substantial portion of suspended matter in the NBP; however their formation in the relatively oxygen-rich waters of the deep Atlantic is considerably more rapid than in the older, oxygen-poor waters of the deep Pacific. Consistent with this explanation, OBS records from ABE dives on the ELSC (SW Pacific) often show relatively low intensity signals in the immediate vicinities of BP stems relative to a few hundred meters away (larger sulfate and sulfide particles responsible for the "smoke" of typical black-smoker type vents tend to fall out during ascent) [37].

[†]National Institute of Advanced Industrial Science and Technology (AIST), Institute for Marine Resources and Environment, Seafloor Environment and Resources Research Group 1, Tsukuba, Ibaraki, 305-8567 Japan

appears to be a reliable indicator of the interception of nascent hydrothermal effluent. For this reason, eH voltage from the sensor is differentiated prior to passing through the latter stages of Algorithm 8. To avoid enhancing quantization noise, the raw sensor output is pre-filtered before applying an approximate numerical derivative. Use of the derivative means weak gradients in redox potential will not be resolved, whereas the steep gradients encountered, for instance, upon entering a BP will be emphasized.

Vertical Velocity Anomaly (VVA) Vertical velocity anomaly (VVA) refers to the vertical component of measured water current velocity. Significant vertical water velocities provide an essentially unequivocal indicator of contact with a BP by definition. Though VVA is a scalar it has no meaningful associated measure of concentration and is consequently not a tracer in the same sense as the other tracers considered. Nevertheless, in practical terms it is analogous to a non-conservative tracer with effectively zero NBP residence time.¹⁶ On account of their specificity, VVA detections are valuable, but also rare. This thesis presents results from two different approaches to measuring VVA aboard ABE: (1) directly, using an on-board ADCP;¹⁷ (2) by comparing vehicle behavior in the vertical plane to a reference model.

Measuring vertical velocity from an on-board ADCP is conceptually straightforward. Within range of the bottom (200 m to 300 m height with ABE's 300 kHz unit), ground-referenced vehicle velocity is simply subtracted from water-referenced vehicle velocity, both measurements being provided by the unit. At greater heights off bottom, differentiated depth provides an accurate, low-noise estimate of vertical vehicle velocity.¹⁸ Potentially complicating factors are that BP interceptions tend to be brief, so that available averaging time is short and may not occupy all beams of the instrument, which can lead to large errors in the computed velocity and possible rejection by the instrument.

Alternatively, a unique methodology developed by Yoerger et al. [135] estimates ambient vertical velocity by comparing the output of a simple one degree-of-freedom model for the vertical dynamics of the vehicle driven by commanded thrust to the

¹⁶Significant vertical water velocities are uncommon within the NBP when not directly associated with BPs.

¹⁷RD Instruments, Inc. 300 kHz Workhorse ADCP/DVL, <http://www.rdinstruments.com/>

¹⁸No such nearly ideal reference exists for horizontal vehicle velocity (cf. § 4.4).

vehicle’s actual, measured depth.¹⁹ The resulting estimate of vertical water velocity represents, to the accuracy of the model, the actual water velocity filtered through the vehicle dynamics [135]. The simplicity of the vehicle model limits its use to approximately straight and level flight at constant speed. Phase-1 dives in the NBP satisfy these criteria and the method resolves vertical velocities of ~ 1.5 cm/s reliably after the removal of turns. Though the detection limit suffers, the method still provides indication of much stronger vertical velocities on 50 m height Phase-2 dives, despite active depth changes by the vehicle while bottom-following.

4.3.4 Background Profile

AUVs are well-suited to collecting data at constant depth;²⁰ however, depth is not ideal as an independent reference variable for defining background profiles. The relative quiescence of the deep ocean and stable stratification thereof make potential density σ , defined relative to some convenient reference pressure, a better choice. This reflects the fact that oceanic mixing occurs primarily along surfaces of constant potential density, which in turn reflects the suppression of vertical mixing by stable stratification (e.g. [57]). Furthermore, anomalies defined with respect to potential density are immune to the vertical displacement of isopycnals due, for instance, to passing internal waves. Potential density, rather than in situ density, removes the effects of pressure on density due to the compressibility of water, which has nothing to do with mixing.

Over the short depth interval of interest to hydrothermal studies, assuming a linear dependence on σ often provides a good fit, and one sufficient to identify the deviations associated with a NBP within a few kilometers of the source.²¹ Under the assumption of a linear dependence on potential density, background profiles will

¹⁹Yoerger’s original model included a static buoyancy force; however, to avoid re-calibrating the model for each dive to account for a changed payload or different survey depth, I have removed this element of the model in favor of band-passing the difference between the measured depth and model output.

²⁰More accurately, AUVs typically fly along isobaric surfaces since depth is usually mapped one-to-one from measured pressure.

²¹Assuming a linear dependence is, however, insufficiently accurate for certain sensitive heat flux computations [120].

satisfy

$$\Delta_{\sigma}v = \frac{\partial v}{\partial \sigma}(\sigma - \sigma_{ref}) + v_{ref} \quad (4.1)$$

where the derivative is presumed constant and (σ_{ref}, v_{ref}) is an arbitrary reference point. It remains to estimate the terms $\frac{\partial v}{\partial \sigma}$ and v_{ref} in (4.1) and to define thresholds for each tracer above which $\Delta_{\sigma}v$ can be reliably regarded as due to hydrothermal input as opposed to background variability or measurement noise.

If available, a CTD cast in nearby waters unaffected by hydrothermal input can provide the parameters of (4.1) for tracers that possess background profiles; however, there are several reasons to prefer a background profile generated using data from the vehicle's own descent:

1. There is no requirement for sensor cross-calibration.
2. The profile could vary significantly between the location or time of a background CTD cast and the survey site and deployment time.
3. The AUV may carry some sensors not also available on the host ship's CTD.
4. Sensors aboard the AUV may have different noise characteristics than identical ones on lowered CTDs (hydrodynamics and horizontal vs. vertical deployment).

The finite battery energy available to an AUV dictates that it be deployed as close as safely possible to the survey location. Consequently the vehicle is likely to pass through the NBP during its descent in which case profile data will be contaminated by the very anomalies we wish to identify. In that case, (4.1) can be used to extrapolate data from immediately above the NBP to greater depth. Results presented subsequently attest to the viability of this approach within the depth intervals corresponding to encountered NBP. Further extrapolation to depths below the NBP tends to be less successful, at least partly for reasons unrelated to the accuracy of (4.1) or to the extrapolation thereof. Persistent hydrothermal activity, particularly in enclosed axial valleys, will alter profiles beneath the NBP relative to uncontaminated off-axis profiles [82], meaning that BP anomalies are more appropriately computed relative to a moving baseline as discussed below.

Identifying depths altered by hydrothermal venting in hydrographic profiles following vehicle recovery is fairly straightforward if somewhat subjective. Of the hydrothermal tracers that are measurable on ABE, OBS has proven the most reliable for

delineating the vertical extent of the NBP. Estimates of the parameters $\frac{\partial v}{\partial \sigma}$ and v_{ref} in (4.1) are then determined by least squares applied to a suitable interval above the top of the NBP. Autonomous implementation of Algorithm 8 would require the vehicle to perform this task independently. I have attained moderate success in this regard with a robust model-fitting technique based on the RANdom SAMple Consensus (RANSAC) algorithm [35]. RANSAC works by randomly selecting a subset of data, fitting the model to this data, counting the number of inliers from the whole dataset based on a threshold criterion, and iterating until converging on the model parameters that maximize the number of inliers. By selecting candidate inliers only from outside of a random interval instead of completely at random, the algorithm can be made to return an estimate of the interval corresponding to the NBP as well as model parameter estimates. The weakest link appears to be selection of the portion of the vehicle’s descent to use as an initial data set. Too much data above the NBP leads to unreliable fits because of nonlinearity in the profile, whereas too little data can lead the algorithm to choose other portions of the descent. A priori approximate knowledge of the depth of the NBP greatly facilitates the choice of a good interval.

Figure 4-5 shows an example of successful automatic NBP interval determination applied to the last 500 m of a descent. Figure 4-6 shows least squares linear fits to the profiles of θ and OBS extrapolated to the seafloor. Removal of background profiles enables data collected from disparate depths to be compared quantitatively and is especially valuable for dives that span multiple depths (e.g., dives 128 and 150 in App. F). Figure 4-7 shows the results of background profile removal applied to OBS and θ time-series data from ABE-126.

Anomalies computed for data from beneath the NBP undergo an additional processing step before declaring detections. Tracer time-series pass through a high pass filter (labeled HPF in Algorithm 8) with a cutoff frequency corresponding to 1 km of trackline at nominal vehicle speed.²² This step removes low frequency trends from non-linearity in the computed profile or from local variations caused by persistent hydrothermal venting and not directly indicative of BP contact. This kind of filtering would be inappropriate for anomaly computation in the NBP where the majority of

²²This choice of cutoff frequency is arbitrary but significantly longer than any vent field yet encountered by ABE while still short on the length scale of an entire ABE dive.

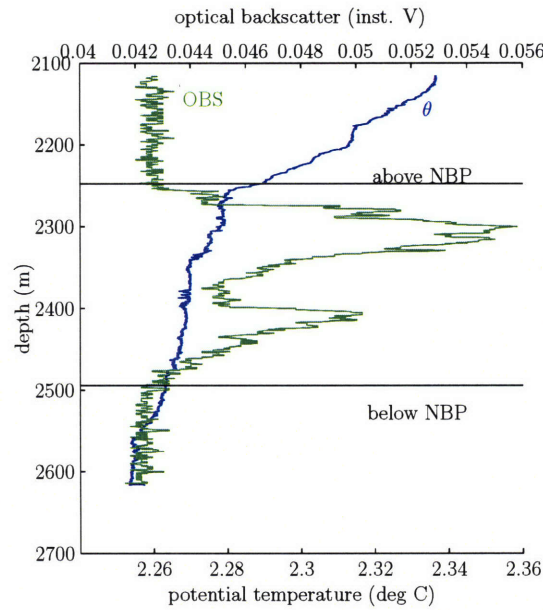


Figure 4-5: Potential temperature and OBS profiles showing contamination by NBP waters and automatic segmentation using OBS of the contaminated portion of the profile. Data are from Phase-1 dive ABE-126.

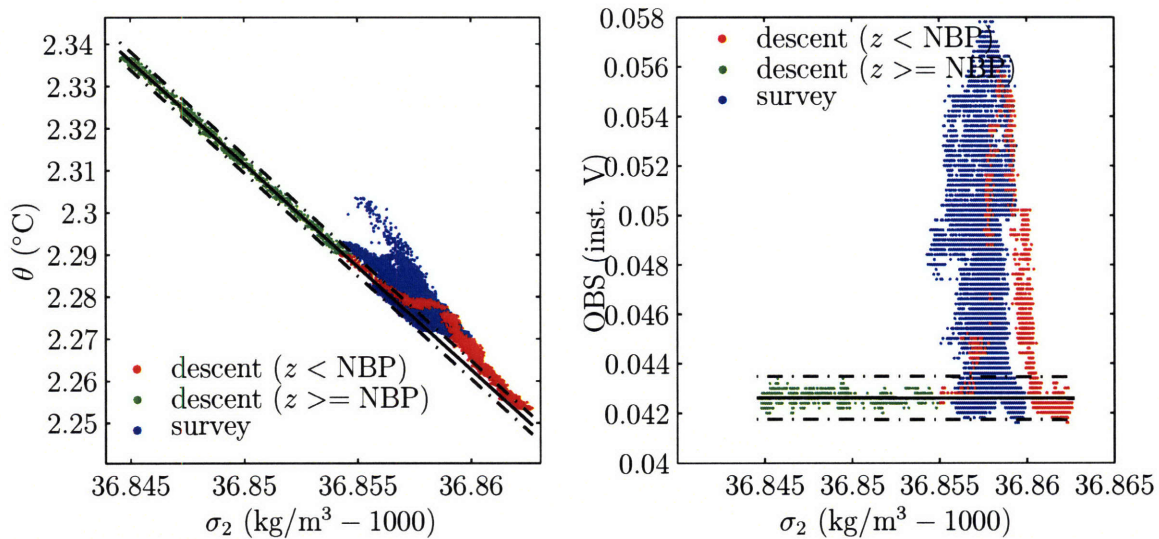


Figure 4-6: Potential temperature and optical backscatter versus potential density for the descent and survey portions of a dive. The solid line is a least squares fit to descent data in an approximately 100 m thick layer immediately above the NBP. Dashed lines indicate the width of the 99.9% prediction interval extrapolated to the approximate potential density at nominal survey depth. Anomalies encountered during the survey computed using these profiles are shown versus horizontal coordinates in Figs. 4-9 & 4-10. Data are from Phase-1 dive ABE-126.

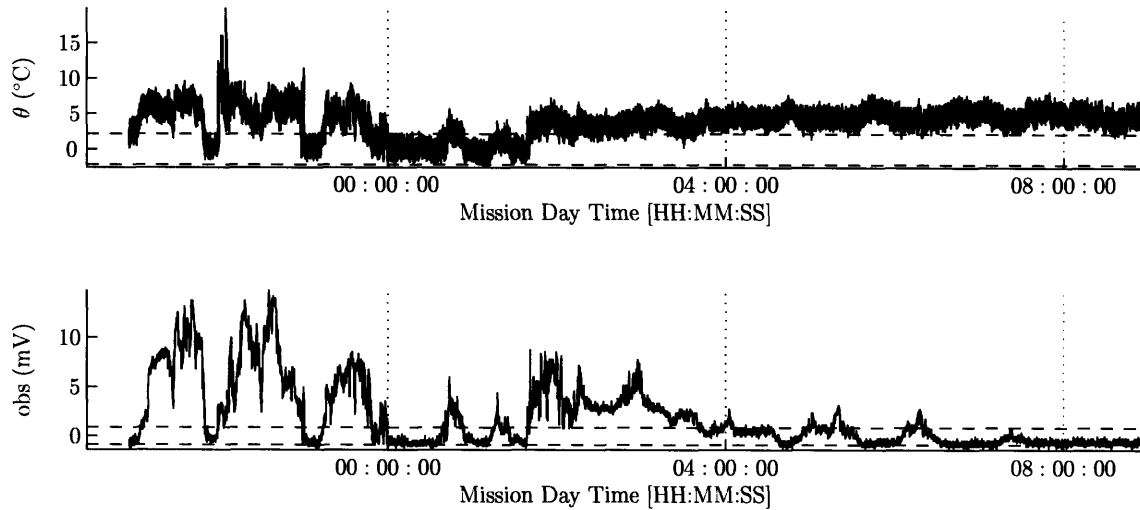


Figure 4-7: OBS and potential temperature anomaly following removal of background profile. Dashed lines indicate the 99.9% prediction interval above which anomalies are declared as indicative of NBP contact. Data are from ABE-126.

a dive can consist of contact with the NBP. Contacts with BP stems, on the other hand, comprise only a small fraction of any given data set. Figure 4-8 shows the results of background profile removal and high-pass filtering applied to all time-series data from ABE-126 (Phase-1).

4.3.5 Declaring Detections

Upon removal of trends, anomalies are classified into either detections or non-detections. Different processing is applied to the declaration of NBP detections than to BP detections. As with the de-trending discussed above, these differences reflect the wide separation in length scale between a NBP and any BP stems feeding it. NBP detections are declared using the standard statistical measure of a prediction interval. Interval width is learned from data known to be uncontaminated collected during the vehicle's descent. Outlier detection methods provide a better fit for BP detection; the method employed here is known as a Hampel Identifier [22].

Ultimately both methods amount to the careful choice of a threshold for each (detrended) anomaly time-series. Their value lies in automatic determination of that threshold value according to a specified probability of false alarm P_F . In both cases, normality of the underlying probability density of anomalies Δv within background

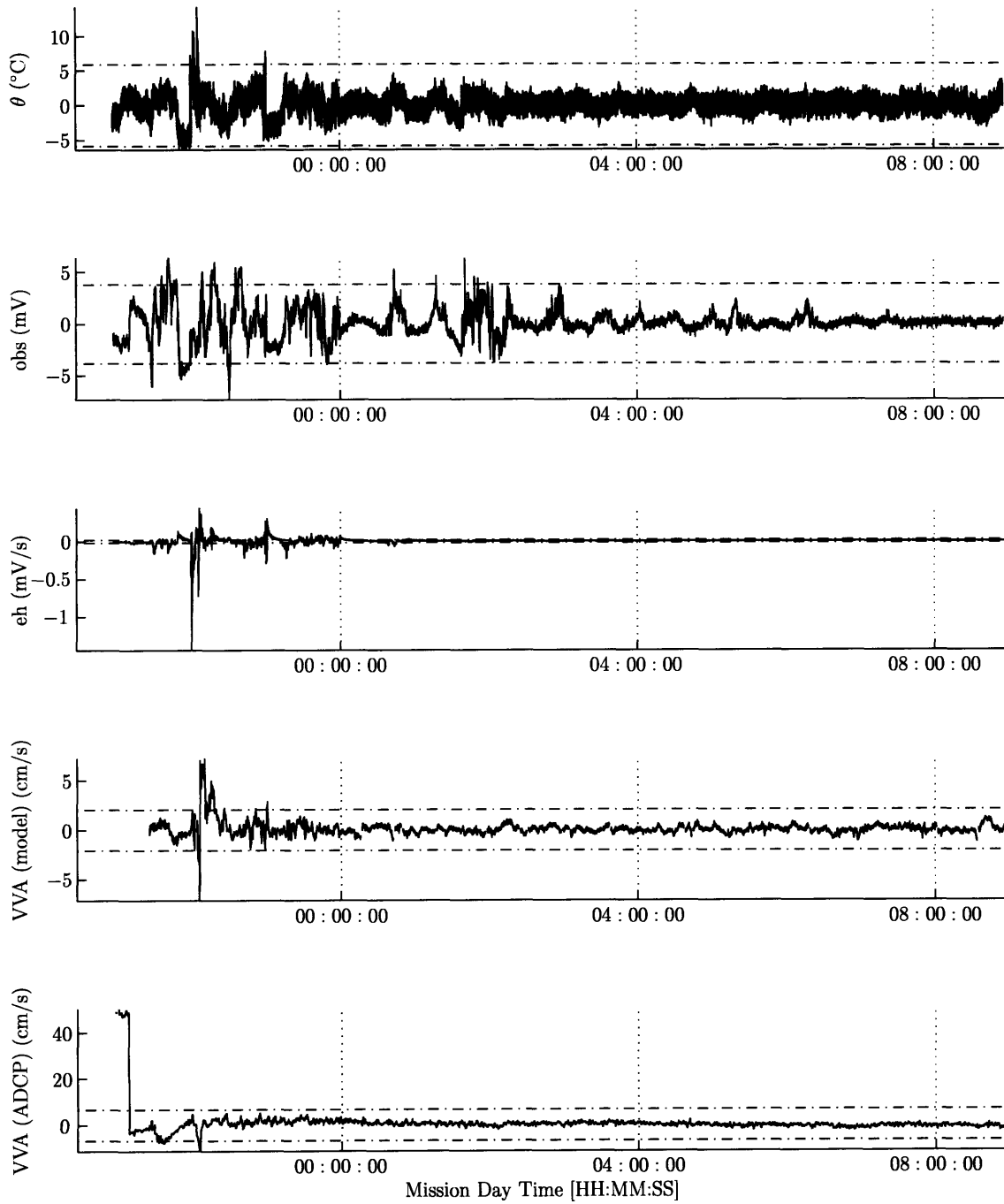


Figure 4-8: Tracer time-series following detrending for buoyant plume detection. Dashed lines indicate thresholds determined by application of a Hampel Identifier to each data set. VVA samples lying outside these thresholds are indicative of BP contact. The other tracers are used as corroborating evidence. Data are from Phase-1 dive ABE-126.

waters is assumed so that given an estimate of the associated variance $\lambda_{\Delta v}$, and assuming successful detrending has removed any mean, P_F is given by:

$$P_F = \int_{|\Delta v| > \phi_v} \mathcal{N}\{u; 0, \lambda_{\Delta v}\} du \quad (4.2)$$

where ϕ_v denotes the threshold for declaring detections. Given a desired P_F the corresponding threshold is

$$\phi_v = (\lambda_{\Delta v})^{1/2} \Phi(1 - P_F/2) , \quad (4.3)$$

where Φ denotes the cumulative distribution function for a standard normal.²³

Unlike classical hypothesis-test based detection theory, the use of prediction intervals and outlier detection avoids the problematic task of specifying a probability distribution for tracer concentration within plumes. However, without this distribution the classical notion of a probability of detection is impossible to compute as is the associated operating characteristic.²⁴ On the other hand, note that a probability of detection so defined refers only to the probability of declaring a detection while immersed in either a NBP or BP. From a survey design perspective, the probability of even encountering a BP is of far greater importance. Key elements of that quantity are developed subsequently in § 4.5 of this chapter; development is completed in Ch. 5.

NBP: Prediction Intervals

Temporal and horizontal variability in the water column, sensor noise, sensor drift, and any nonlinearity in the true profile will all contribute to variation about the estimated linear profile (4.1). Under the assumption that the resulting errors are independent, identically distributed and normally distributed, prediction intervals that specify the probability of future values occurring within some interval can be com-

²³Eq. (4.2) applies to bilateral anomalies for which absolute values exceeding the threshold are regarded as outliers and hence as detections. If, for example, only negative values qualify as outlier candidates then the analogous expression for unilateral anomalies is $P_F = \int_{\Delta v < -\phi_v} \mathcal{N}\{u; 0, \lambda_{\Delta v}\} du$, and the threshold as a function of desired P_F is $\phi_v = \Phi(1 - P_F)$.

²⁴The operating characteristic of a generic detector consists of a curve on which which increased probability of detection can be traded off with increased probability of false alarm by decreasing the detection threshold.

puted using standard methods (e.g., [45], pg. 633). Values that fall outside these prediction intervals are statistically unlikely to have been drawn from the same distribution as background values and are therefore likely to represent NBP contact. Of course, the extent to which the actual data conforms to the assumption of normality will influence the degree to which this is true in practice. This approach has the advantage that all thresholds are defined after descent but prior to commencing the survey portion of a dive, thereby potentially enabling real-time adaptive responses to NBP detection.

Figure 4-6 shows 99.9% prediction intervals ($P_F = 0.001$) for OBS and potential temperature θ along with data from the survey portion of the dive indicating a large fraction of data collected on this dive sampled NBP water. Samples tagged as detections are plotted in Figs. 4-9 & 4-10 and illustrate the utility of multi-tracer data to interpret plume structure. Conservative potential temperature anomalies serve to delineate the extent of the non-buoyant plume and suggest interception of a buoyant plume at the northwest corner of the survey. OBS plays a complementary role with a stronger signal over much of the plumes extent but with less predictable intensity due to the complex nonconservative behavior of the hydrothermal particulates. Differentiated redox potential $d/dt(\text{eH})$ is helpful for identifying the younger parts of a NBP; however, the sensor's dynamics imply gross violations of the assumption of normality above, and better results have been obtained using the outlier detection method discussed next. VVA is specific to BP stems and not considered here.

BP: Hampel Identifier

Outlier detection based on the Hampel Identifier [22, 103, 104] is the key processing step applied to BP detection. The procedure relies on robust estimates for the mean and standard deviation of the supposedly Gaussian background computed using the median and median absolute deviation (MAD). Multiplication by the factor 1.4826 in Algorithm 8 makes the expected value of the result equal to the standard deviation for normally distributed data [104]. Data that exceed the median by a specified number of MAD are declared outliers, or in the present application, detections. As above, this threshold is determined from a desired P_F using the scaled MAD estimate for the variance $\lambda_{\Delta v}$ in (4.3). This method requires batch-processing of the entire record from the survey portion of a dive before declaring BP detections. Moving-window

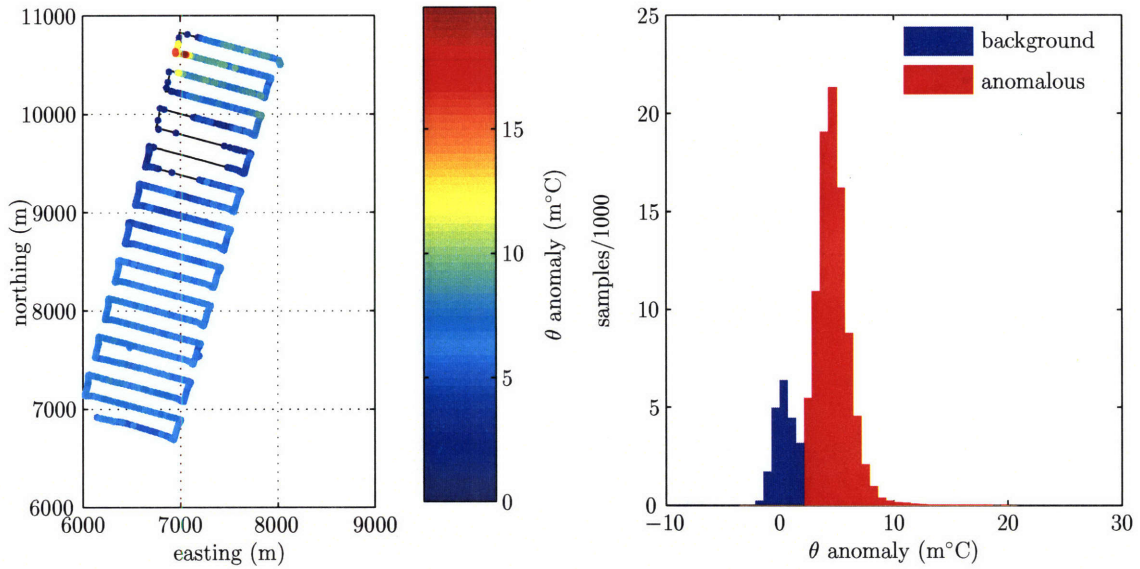


Figure 4-9: Potential temperature anomaly $\Delta_{\sigma}\theta$ observed during Phase-1 dive ABE-126: **(left)** anomalies plotted at the location they were observed; **(right)** the distribution of anomalies observed over the course of the survey. Color indicates intensity. The thin black line in the left plot shows the vehicle track through background regions. The strongest θ anomalies are associated with interception of a buoyant plume near the northwest corner of the survey. With the exception of a small patch of background water to the south of this location, the entire survey appears to suggest contact with hydrothermally influenced water; though anomalies in the southern portion of the survey are relatively weak and may represent gradually varying hydrography.

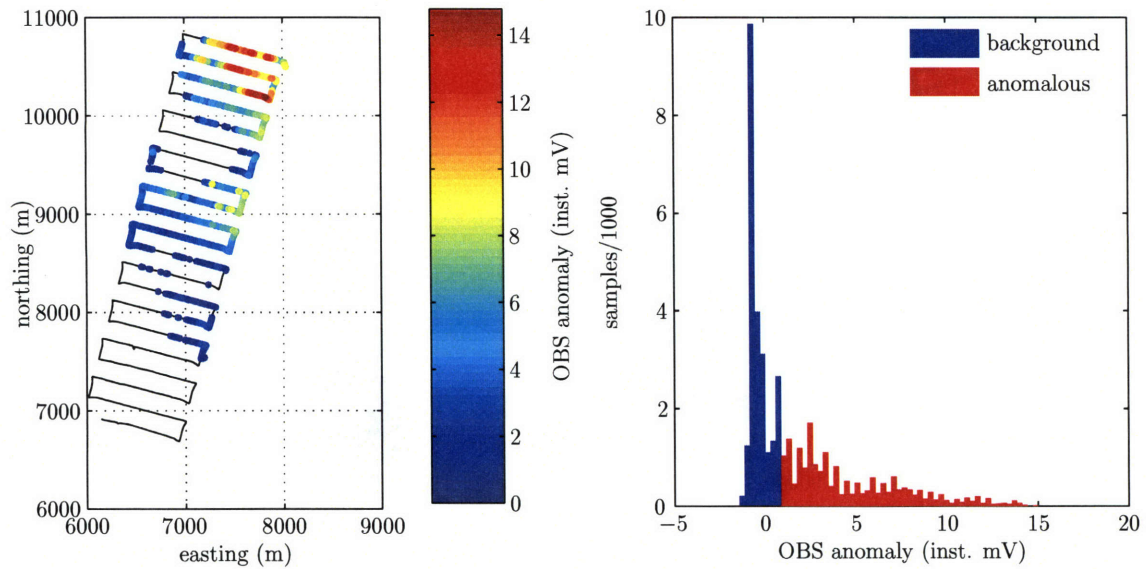


Figure 4-10: Optical backscatter anomaly $\Delta_{\sigma}\text{OBS}$ observed during Phase-1 dive ABE-126: **(left)** anomalies plotted at the location they were observed; **(right)** the distribution of anomalies observed over the course of the survey. The nonconservative nature of OBS is evident. The strongest OBS anomalies were observed in a patch several hundred meters to the west from where the vehicle intercepted buoyant plume water. A second patch of high OBS to the south of the strongest anomalies confirms the hydrothermal origin of weaker $\Delta_{\sigma}\theta$ anomaly evident in Fig. 4-9.

versions of the Hampel Identifier exist [104] but would be inappropriate in this case because minor anomalies in relatively quiescent portions of dives would be identified as detections despite being much smaller in magnitude than anomalies associated with actual BP contact, which tend to occur in backgrounds with greater variability.

Figure 4-11 shows the locations of outliers for each tracer from a sample Phase-1 dive in the NBP. Within the NBP, these plots complement those of Figs. 4-9 & 4-10. Both potential temperature and redox potential (eH) suggest interception of nascent plume water near the northwest corner of the survey. The latter also reveals the relative youth of the NBP encountered over the northern portion of the survey. The single group of VVA anomalies confirms interception of a BP at the northwest corner of the survey, and the lack of other VVA anomalies indicate that the remaining eH anomalies were probably not associated directly with BP activity. Examples of this processing applied to Phase-2 dives are in App. F. All tracers and dives were processed with $P_F = 10^{-8}$, corresponding to declaring outliers above approximately six standard deviations from zero.

BP: Consensus Detection Declaration

Applied to individual anomaly time-series, the Hampel Identifier does a qualitatively good job of picking out data associated with nearby BP activity. However, many more OBS, eH, and θ anomalies tend to be identified as outliers than VVA anomalies. Regardless of the specific tracers employed, the logical expression of the last block in Algorithm 8 represents one means of deciding whether, taken together, the anomalies indicate a BP detection or contradict one another, suggesting some detections may have an alternate explanation.²⁵

In part, the rarity of VVA anomalies reflects a relatively high noise level; nonetheless, I will take the perspective that only significant VVA qualifies as indicative of definitive BP contact. To improve rejection of spurious VVA outliers, the final step in Algorithm 8 requires that VVA detections be accompanied by eH detections. In

²⁵Decisions based on the magnitudes of anomalies represent another, but more complex approach, as the required decision rule would consist of a mapping $f : \mathbb{R}^k \rightarrow \mathbb{B}$ instead of a relatively simple boolean expression $f : \mathbb{B}^k \rightarrow \mathbb{B}$.

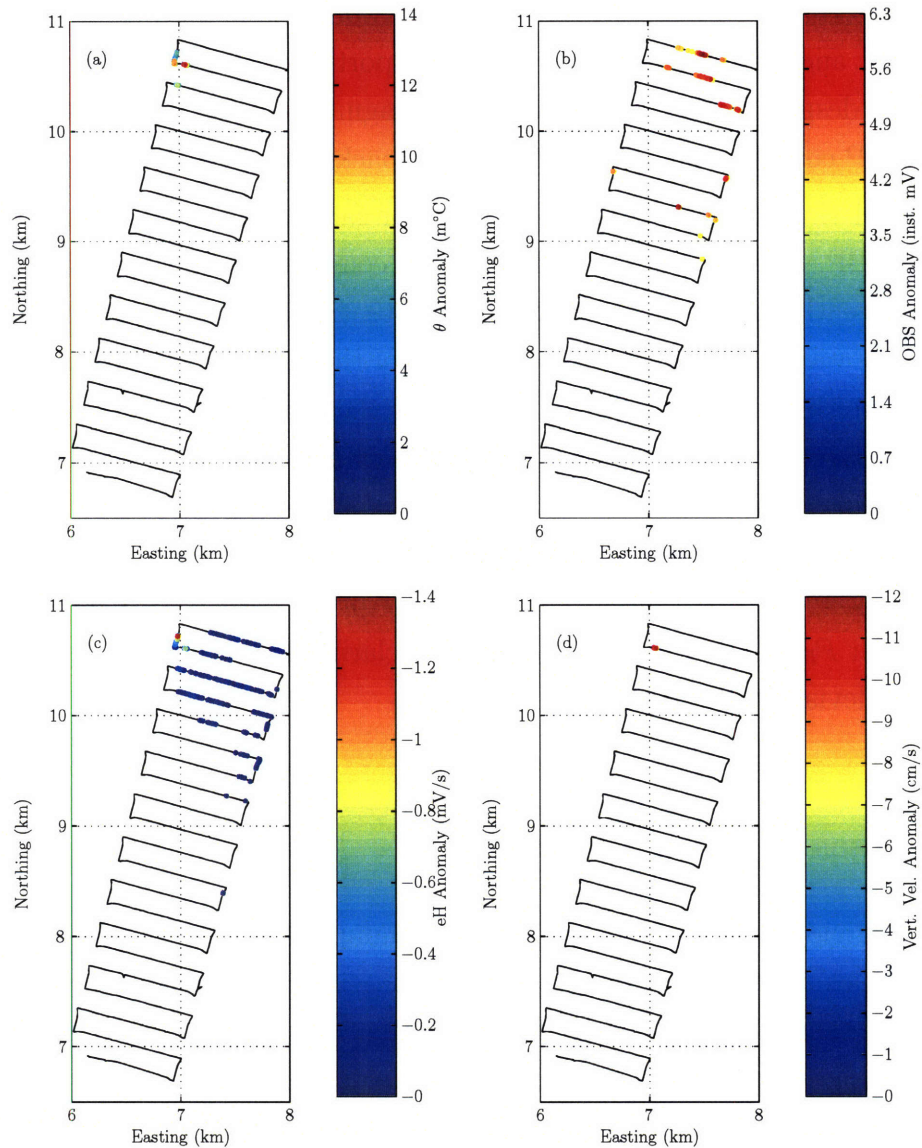


Figure 4-11: Anomalies classified by a Hampel Identifier as BP detections during Phase-1 dive ABE-126: **(a)** potential temperature θ ; **(b)** optical backscatter (OBS); **(c)** redox potential (eH); **(d)** vertical velocity anomaly (VVA). Color indicates intensity. VVA and θ both indicate interception of a BP near the northwest corner of the survey. Redox potential decreased rapidly somewhat before this encounter, but additional spikes correlate with the main detection. In addition, eH shows plume waters immediately to the east of the detection location to be composed of nascent plume water consistent with the downwind location of these detections (cf. App. F). OBS anomalies were considerably more intense within this nascent portion of the NBP than in the BP encountered, probably reflecting the slow formation of iron oxides in the oxygen-poor waters of the deep Pacific [37]. VVA anomaly intensity is shown according to measurements from ABE's on-board ADCP; however, problems with the instrument during this dive necessitated that only those outliers also associated with model-based VVA outliers be considered as candidates for detections.

terms of the boolean function f in Algorithm 8,

$$\delta^t = [\delta_{VVA}^t \cap \delta_{eH}^t] \quad (4.4)$$

where the $\delta^t t$ denote binary indicator variables associated with each tracer at time t , and the square brackets denote a logical operation that produces a boolean result. All buoyant plume detections used subsequently in § 4.5 and Ch. 5 were declared according to this rule. This choice for f reflects the thinking that the nascent, as yet unoxidized plume water within a BP should be accompanied by a strong drop in eH relative to background [95]. Potential temperature could serve a similar function without concern for sensor response characteristics; however, possible changes in sign of the anomaly make θ somewhat less attractive. The presence of OBS anomalies, while reliable indicators of NBP contact, appear often enough outside BP stems,²⁶ or within them at relatively low intensities,²⁷ to warrant dismissal in favor of one or both of eH and θ .

4.3.6 Results from Additional Dives

The Phase-1 dive discussed in this section (ABE-126) was selected because interpretation of the tracer records was straightforward and groundtruth locations of vent sites were well-resolved from subsequent dives on the site. Anomaly maps for all other dives discussed in this work are provided in App. F. These include both Phase-1 and Phase-2 dives; however, not all of these included BP detections according to the methodology developed above. For these dives, successful vent localization was still sometimes achieved based on manual interpretation of the data. The maps presented here and in App. F suggest that proximal BP activity is sometimes associated with elevated anomalies in the other tracers even in the absence of any significant VVA anomaly. That appears to be especially true of redox potential (eH) within the NBP, which is unique among the tracers measured on ABE in its ability to distinguish nascent, though not necessarily still buoyant, plume water. The last component of the methodology above could therefore be criticized on grounds of ignoring this poten-

²⁶Possible explanations include particle fallout from above and sediment resuspension.

²⁷This effect may reflect insufficient water-column residence time for precipitates to have been produced or variations in source and background water chemistry.

tially valuable source of data. Subsequent development in § 4.5 and Ch. 5 capitalizes on the availability of models that describe BP evolution and hence provide a tool with which to predict source location on the seafloor following a BP detection. To fully capitalize on eH detections within the NBP will require the development of an analogous model capable of predicting the locations of these detections relative to the locations of buoyant feeder plumes.

4.4 Water Velocity Profiles from On-Board ADCP

In principle, knowledge of ambient crossflow current should aid in constraining the source of hydrothermal plumes. Because water current velocity cannot be observed directly from a moving platform, the essential element common to all methods for estimating water current velocity from moving platforms is vector-subtraction of the platform’s velocity-through-water (VTW) from the platform’s velocity-over-ground (VOG).²⁸ Simply because of the water’s proximity to the vehicle, it is relatively easy to measure velocity-through-water (VTW). In contrast, measuring VOG can be much more difficult because the ground, or some other suitable reference is often far from the vehicle and outside the range of on-board sensors. If available, differentiated position measurements can be substituted for direct measurements of VOG. As the following will show, crossflow determination using the latter can be sensitive to small orientation errors, and every effort must be made to remove such errors. This section concludes by presenting crossflow velocity measured both with and without the availability of an instrument-frame VOG measurement. Success is attained in either case, albeit with reduced temporal resolution in the latter.

4.4.1 Error Sources

To compute earth-referenced water current velocity, three types of measurements are required: (1) vehicle VTW, $\mathbf{u}_w \in \mathbb{R}^3$; (2) vehicle VOG, $\mathbf{u}_g \in \mathbb{R}^3$; (3) vehicle attitude (roll, pitch, heading). In underwater vehicle applications using a combined ADCP and Doppler velocity log (DVL) to provide velocity measurements, the dominant error sources consist of approximately zero-mean velocity noise and instrument (or vehicle) orientation biases. With the exception of heading, underwater vehicle attitude is relatively easy to measure with low-cost tilt-meters on account of the ubiquitous gravity vector and relatively slow accelerations involved in vehicle maneuvers. Heading can be measured cost-effectively with magnetic compasses, however careful calibration to account for the magnetic properties of the platform is essential [14]. Imperfect calibration produces so-called hard-iron and soft-iron errors, which result in heading bias that is a function of true heading (heading-dependent heading bias) [14]. Even

²⁸I will reserve the more common terms speed-over-ground (SOG) and speed-through-water (STW) to refer to scalar speeds, using VOG and VTW to denote vectors.

with proper calibration, environmental variability of the Earth’s magnetic field can produce local biases that cannot be calibrated away a priori. More costly fiber-optic north-seeking gyroscopes provide nearly ideal attitude measurements (0.1° accuracy); these have enabled Kinsey and Whitcomb [56] to solve the problem of in situ relative alignment between instruments for DVL-equipped vehicles operating near bottom.²⁹ However, their method does not address the heading-dependent heading bias as results from use of a magnetic compass and requires that the Doppler sensor be within range of the seafloor to measure VOG.

The key difference between zero-mean velocity noise and orientation or attitude bias (systematic error) is that the former can be attenuated by averaging. Tidal currents vary slowly (hrs.) whereas typical update rates for navigation sensors on AUVs are $O(0.1 - 10 \text{ Hz})$ [128] so that averaging readily attenuates zero-mean noise, assuming, of course, that the averaging can be applied in the earth-frame. Attitude and orientation bias corrupt the transformation of instrument-frame velocities into the earth frame. If both VOG and VTW are measured in the instrument-frame, the error incurred for small biases is commensurately small; however, as the ensuing analysis shows, even small biases can completely obscure estimated water current velocity when VOG cannot be measured natively in the same frame as VTW.

4.4.2 Effect of Heading Error

Since many ADCPs suitable for use on AUVs include tilt-meters for measuring pitch and roll of the unit, I will consider only the effect of a heading error. For simplicity, consider a vehicle translating along a straight line at constant attitude (roll, pitch, heading) in a steady current \mathbf{w} collecting noiseless pitch, roll, VTW \mathbf{u}_w , and VOG \mathbf{u}_g measurements, but where either heading ψ , static vehicle-relative velocity sensor orientation ${}^v_i R$, or both are subject to unknown bias. Water velocity is computed from these measurements in different ways depending on whether \mathbf{u}_g is measured in

²⁹Kinsey and Whitcomb [56] report a method for in-situ calibration of attitude and acoustic Doppler sensors. Their method solves for a static sensor orientation that minimizes the difference between dead-reckoned position estimates from the Doppler sensor and absolute position measurements from a long baseline (LBL) system.

the instrument-frame (4.5a) or earth-fixed reference frame (4.5b):

$${}^e\mathbf{w} = {}^eR(\psi) {}^vR({}^i\mathbf{u}_g - {}^i\mathbf{u}_w) \quad \text{inst.-frame;} \quad (4.5a)$$

$${}^e\mathbf{w} = {}^e\mathbf{u}_g - {}^eR(\psi) {}^vR({}^i\mathbf{u}_w) \quad \text{earth-frame.} \quad (4.5b)$$

Frame-notation is described in App. A; underscored frame identifiers denote the leveled versions of those frames. The transformation between the leveled instrument frame \underline{i} and leveled vehicle frame \underline{v} can be expressed in terms of the unleveled instrument orientation rotation matrix vR as $\frac{v}{i}R = \frac{v}{v}R \frac{v}{i}R \frac{i}{i}R$. For constant vehicle pitch and roll, $\frac{v}{i}R$ describes a static rotation about the gravity vector. Under these conditions, a error in vR results in a static bias indistinguishable from vehicle heading error.

Figure 4-12 depicts the effect of heading error (and sensor orientation error) on the horizontal earth-frame water velocity estimate assuming otherwise noiseless measurements. When both \mathbf{u}_g and \mathbf{u}_w are measured in the instrument-frame, the magnitude of the water current velocity $w \triangleq |\mathbf{w}|$ is unaffected, whereas its direction $\angle w \triangleq \angle \mathbf{w}$ is rotated from true by exactly the heading error $\tilde{\psi}$. Changes in vehicle attitude will produce dynamic $\angle w$ errors and hence affect the earth-frame averaging of \mathbf{w} ; however, since $\tilde{\psi}$ is typically small and most AUV missions consist of long, straight tracklines, good estimates of \mathbf{w} are generally obtainable. Alternately, when an instrument-frame VOG estimate is unavailable and replaced by an earth-frame estimate, heading bias incurs errors in both water velocity magnitude w and direction $\angle w$. The difference in sensitivity to heading-bias between \mathbf{w} computed via (4.5a) versus (4.5b) can be understood intuitively by considering the vector subtraction depicted in Fig. 4-12. In case (b), the length of two sides of the triangle and the angle between them are known, hence the triangle is completely constrained except for its orientation in the earth-fixed reference frame. In case (b), the length two sides of the triangle are known, but the angle between them is subject to error, hence the third side (\mathbf{w}) is subject to both magnitude and direction error.

Whereas case (b) in Figure 4-12 is straightforward, the effect of heading bias in case (c) is dependent on both crabbing angle γ (defined as the angle between \mathbf{u}_g and \mathbf{u}_w) and on the relative magnitudes of \mathbf{u}_g and \mathbf{u}_w . Henceforth, I restrict my attention to case (c): VOG measured in an earth-fixed reference frame. Figure 4-13 depicts

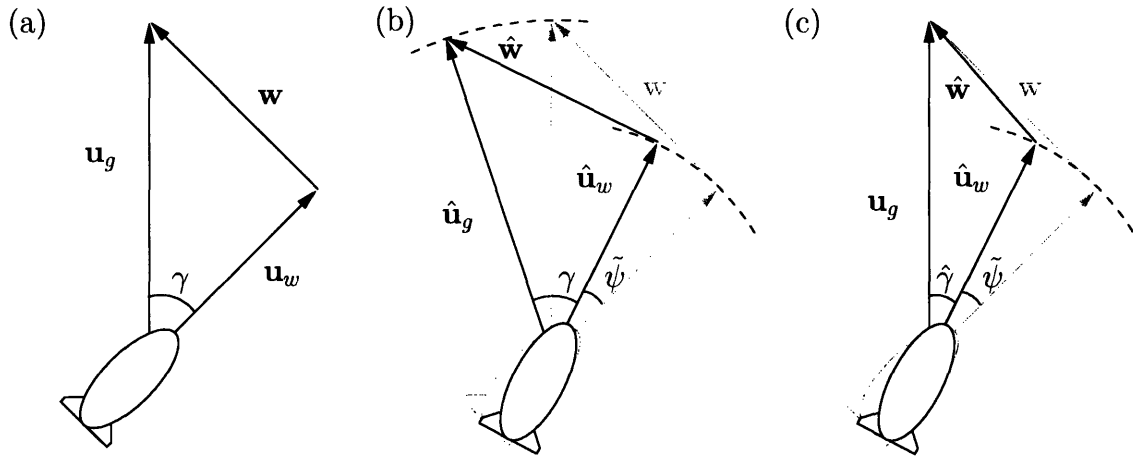


Figure 4-12: Effect of heading error on earth-frame water velocity estimates: **(a)** no heading error ($\tilde{\psi} = 0$); **(b)** non-zero heading error with both VOG \mathbf{u}_g and VTW \mathbf{u}_w measured in the instrument-frame; **(c)** non-zero heading error with \mathbf{u}_g measured in the earth-frame and \mathbf{u}_w measured in the instrument-frame. The crabbing angle is denoted by γ . When both \mathbf{u}_g and \mathbf{u}_w are measured in the frame of the ADCP, $\tilde{\psi} \neq 0$ results in a rotated water velocity estimate, but no magnitude error. If the instrument-frame VOG measurement is unavailable from the ADCP (e.g. no bottom-lock), then an earth-frame VOG estimate from LBL navigation can be substituted; however, $\tilde{\psi} \neq 0$ perturbs both the magnitude and direction estimate of \mathbf{w} . The resulting error is dependent both on γ and the relative magnitudes of \mathbf{u}_g and \mathbf{u}_w .

the dependence of water current velocity \mathbf{w} on γ for constant STW and two SOGs corresponding to a tailwind ($|\mathbf{u}_g| > |\mathbf{u}_w|$) and headwind ($|\mathbf{u}_g| < |\mathbf{u}_w|$).³⁰ The water current velocity magnitude w is attained by application of the law of cosines to the triangle depicted in case (c) of Fig. 4-12:

$$|\mathbf{w}|^2 = |\mathbf{u}_g|^2 + |\mathbf{u}_w|^2 - 2|\mathbf{u}_g|^2|\mathbf{u}_w|^2 \cos \gamma . \quad (4.6)$$

Its direction $\angle w$ is given by the law of sines:

$$\frac{\sin \angle w}{|\mathbf{u}_w|} = \frac{\sin(\pi - \angle w - \gamma)}{|\mathbf{u}_g|} . \quad (4.7)$$

Non-dimensionalization of SOG and water current speed w by vehicle STW yields the non-dimensional variables $u'_g \triangleq |\mathbf{u}_g|/|\mathbf{u}_w|$ and $w' \triangleq |\mathbf{w}|/|\mathbf{u}_w|$. Equations (4.6) and (4.7) become:

$$\begin{aligned} w' &= \left(u_g'^2 + 1 - 2u_g' \cos \gamma \right)^{\frac{1}{2}} \\ \angle w &= \tan^{-1} \left(\frac{\sin \gamma}{u_g' - \cos \gamma} \right) . \end{aligned} \quad (4.8)$$

When VOG is measured in an earth-fixed coordinate frame, heading error is equivalent to an error in crabbing angle γ . Therefore, the non-dimensional magnitude error \tilde{w}' and direction error $\angle \tilde{w}$ due to an error in heading $\tilde{\psi}$ are:

$$\begin{aligned} \tilde{w}'(\tilde{\psi}) &= w'(\gamma + \tilde{\psi}) - w'(\gamma) \\ \angle \tilde{w}(\tilde{\psi}) &= \angle w(\gamma + \tilde{\psi}) - \angle w(\gamma) \end{aligned} \quad (4.9)$$

A Taylor expansion of these expressions about $\tilde{\gamma} = 0$ yields

$$\begin{aligned} \tilde{w}'(\tilde{\psi}) &= \frac{u_g' \sin \gamma}{(u_g'^2 + 1 - 2u_g' \cos \gamma)^{\frac{1}{2}}} \tilde{\psi} + O(\tilde{\gamma}^2) \\ \angle \tilde{w}(\tilde{\psi}) &= \frac{u_g' \cos \gamma - 1}{u_g'^2 + 1 - 2u_g' \cos \gamma} \tilde{\psi} + O(\tilde{\gamma}^2) . \end{aligned} \quad (4.10)$$

³⁰These definitions of headwind and tailwind assume that vehicle speed exceeds water current speed ($|\mathbf{u}_w| > |\mathbf{w}|$), a necessary condition for safe vehicle operation.

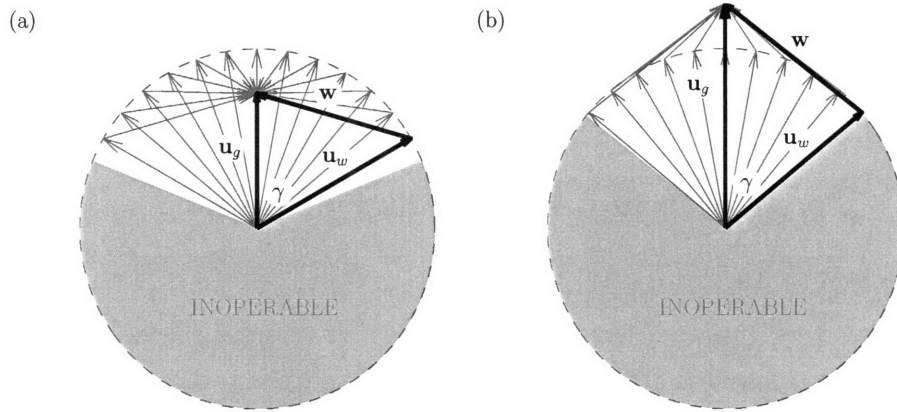


Figure 4-13: Schematic dependence of water velocity measurement \mathbf{w} on crabbing angle γ for two earth-frame SOG $u_g = |\mathbf{u}_g|$ to STW $u_w = |\mathbf{u}_w|$ ratios u_g/u_w : **(a)** headwind with $u_g/u_w = 0.75$; **(b)** tailwind with $u_g/u_w = 1.25$. Heading error is additive to crabbing angle γ and produces a false crabbing angle estimate. These plots indicate the resulting error in the \mathbf{w} estimate is strongly dependent on the actual crabbing angle γ . To see this, consider that for a given γ , the variation in the magnitude and direction of \mathbf{w} for small $\tilde{\psi}$ can be assessed by altering \mathbf{w} such that its tail lies on a line tangent to the dashed circle at the actual γ . For instance, near $\gamma = 0$, heading errors always result in small current magnitude errors, but large directional errors. Large crabbing angles produce magnitude errors that are much more severe in a tailwind than in a headwind, whereas directional errors are more severe in a headwind. The grayed out portions of each pie represent unsafe operating conditions where current speed exceeds vehicle speed ($|\mathbf{w}| > u_w$).

Since heading errors should be small, the leading terms in these expressions approximate water current velocity error per unit heading error. They are plotted as functions of u'_g for various γ in Fig. 4-14.

The key insight offered by Fig. 4-14 is that water current velocity estimation for small crabbing angles $\gamma \approx 0$ and little difference between SOG and STW $u'_g \approx 1$ is sensitive to heading error. These conditions imply either low current magnitudes ($w \approx 0$), or else a “fast vehicle” ($|\mathbf{u}_w| \gg |\mathbf{w}|$). Grossly incorrect $\angle w$ may be irrelevant for $|\mathbf{w}| \approx 0$; indeed, $\angle w$ is undefined for $|\mathbf{w}| = 0$. However, current velocity magnitude error due to heading bias is potentially severe for a fast vehicle because the dimensional error is proportional to vehicle speed.

In terms of non-dimensional water current speed w' , a fast vehicle corresponds to $w' \rightarrow 0$, or equivalently to $\gamma \rightarrow 0$ and $u'_g \rightarrow 1$. In the limit, the leading terms of

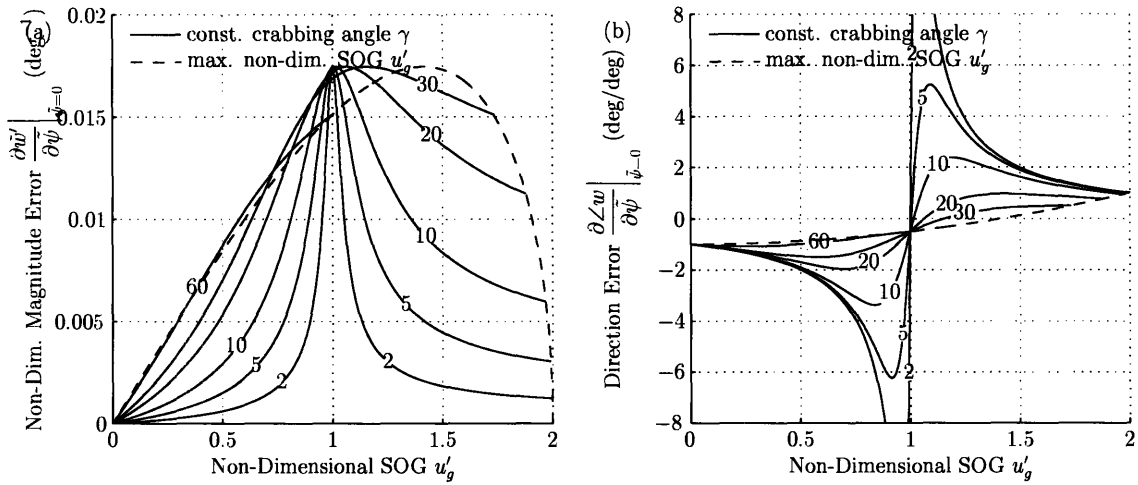


Figure 4-14: Non-dimensional error in water velocity magnitude and direction from heading error as a function of non-dimensional SOG $u'_g = |\mathbf{u}_g|/|\mathbf{u}_w|$ for various crabbing angle γ : (a) first-order magnitude error per unit heading error $\tilde{\psi}$; (b) first-order direction error per unit heading error $\tilde{\psi}$. Over the range of u'_g plotted, heading errors of $\psi = 1^\circ$ produce water velocity magnitude errors of between approximately 1% and 2% of vehicle speed. Faster vehicles and slower currents make water velocity magnitude progressively more sensitive to heading error. The same is true for water velocity direction: slow currents and relatively fast vehicles imply small crabbing angles $\gamma \approx 0$ and $u'_g \approx 1$. These conditions produce progressively larger direction errors per unit heading error as $u'_g \rightarrow 1$ and $\gamma \rightarrow 0$ until attaining a singularity at $u'_g = 1$. The dashed lines in both plots indicate the maximum crabbing angle for a given u'_g under the constraint that vehicle speed exceed current speed ($|\mathbf{w}| > |\mathbf{u}_w|$) to allow for safe vehicle operation.

(4.10) are singular. To avoid this uninformative singularity, consider the behavior of the water current magnitude and direction errors in its vicinity, that is, as $u'_g \rightarrow 1$ for small $\gamma \approx 0$:

$$\begin{aligned}\lim_{w' \rightarrow 0^+} \tilde{w}' &= \lim_{u'_g \rightarrow 1} \frac{u'_g \gamma}{u'_g - 1} \tilde{\psi} \\ \lim_{w' \rightarrow 0^+} \angle \tilde{w} &= \lim_{u'_g \rightarrow 1} \frac{1}{u'_g - 1} \tilde{\psi} .\end{aligned}\tag{4.11}$$

Evidently both types of water current velocity error become extremely sensitive to heading error $\tilde{\psi}$ for fast vehicles when VOG cannot be measured in the same frame as VTW.

4.4.3 Results

ABE is equipped with a 300 kHz ADCP/DVL³¹ and a calibrated magnetic flux gate magnetometer³² with 3-axis magnetic flux density and tilt-compensated heading outputs.³³ In light of the preceding discussion, every effort has been made to eliminate heading bias and errors in the static orientation of the ADCP on the vehicle:

- ABE executes a heading calibration spin during its descent, the data from which is used to determine any remaining hard-iron error using a technique based on finding a hard-iron magnetization vector that fits leveled magnetic flux density to a circle for each complete vehicle revolution [14].
- A static, leveled (single parameter) instrument orientation is computed by comparing VOG from bottom-track DVL velocity to VOG from differentiated LBL position. This compensates for sensor misalignment at a static vehicle pitch and roll and is adequate for a passively stable vehicle like ABE.

³¹RD Instruments, Inc. Navigator ADCP, rdinstruments.com.

³²PNI Corp. TCM2 electronic compass module, www.pnicorp.com

³³ABE's ADCP contains its own heading reference; however, the unit has not yet been calibrated for the magnetic properties of the vehicle owing to the awkwardness of precisely orienting a 700 kg vehicle.

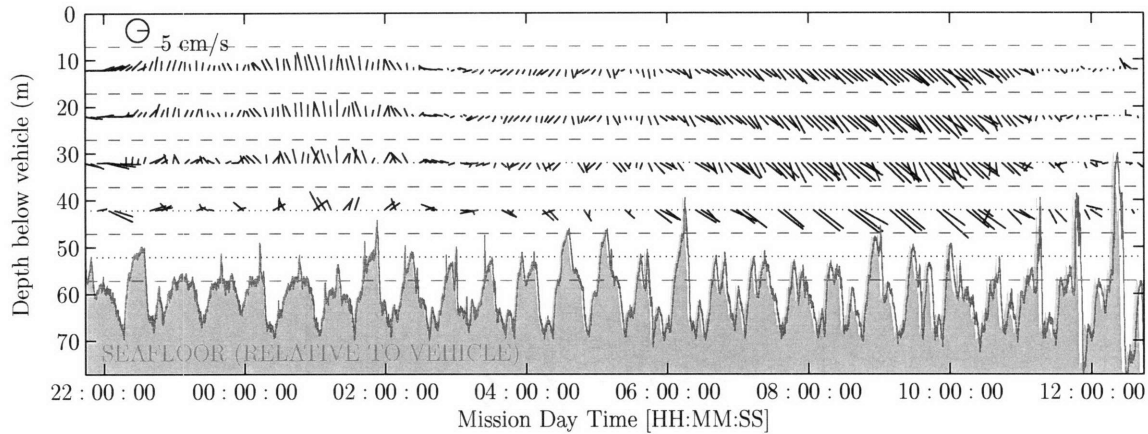


Figure 4-15: Crossflow speed and direction profiles vs. time (5 min averages) from Phase-2 dive ABE-151. Dashed lines indicate the boundaries between 10 m bins beneath the vehicle. There appears to be little variation of crossflow speed or direction with depth compared to temporal variability, though for this type of plot, spatial variability will contribute to the apparent temporal variability. Gaps indicate interference from the seafloor. Bottom lock was maintained throughout the duration of this 50 m height dive.

The next two sections compare results for a water current velocity record acquired by ABE when using bottom-track velocity from ABE’s DVL for VOG versus differentiated position estimates from LBL acoustic navigation.

Instrument-Frame VOG

ABE’s ADCP/DVL is capable of acquiring bottom-lock between heights of 200 m and 300 m above the seafloor. All Phase-2 dives fall within this range, and bottom-lock was available on some Phase-1 dives as well. Figure 4-15 shows profiles of horizontal crossflow averaged over 5 min intervals observed during ABE-151, a 50 m height Phase-2 dive at SMAR. The impressively variable crossflow magnitudes and directions observed are also shown in Fig. 4-16 from a birds-eye perspective. Figure 4-16 in particular illustrates a temporal and spatial variability characteristic of the records acquired by ABE (cf. App. F).

Earth-Frame VOG

ABE is a relatively slow AUV; nonetheless, referring to Fig. 4-14, a heading bias of just 1° in weak crossflows will produce approximately 1 cm/s worth of crossflow speed error

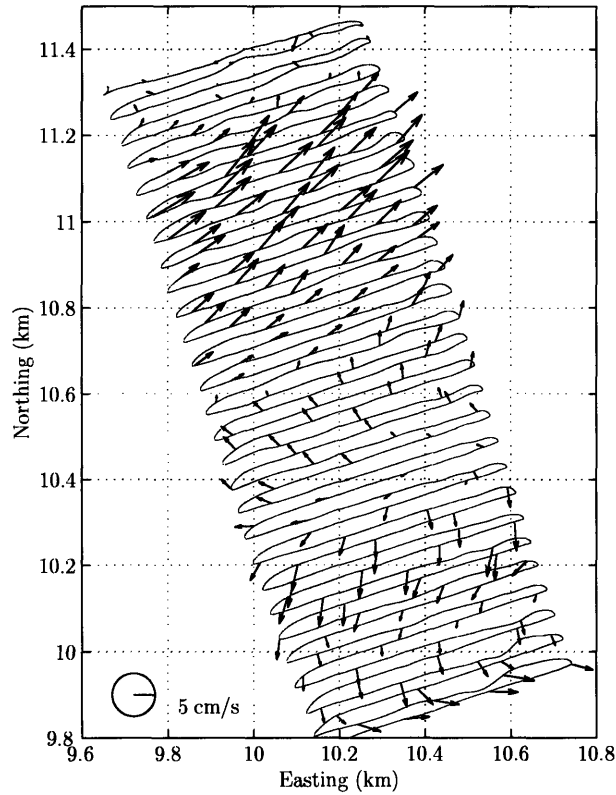


Figure 4-16: Birds-eye view of crossflow speed and direction observed during ABE-151 (5 min averages). Spatial variability in crossflows is especially evident on tracklines near coordinates (10.4,10.4). The eastern ends of these tracklines indicate almost no flow, whereas a westerly component appears persistent on their western ends despite a complete reversal of the northerly component of the flow as the vehicle progressed north. The tracklines shown span about 14 h. Bottom lock was maintained throughout the duration of this 50 m height dive.

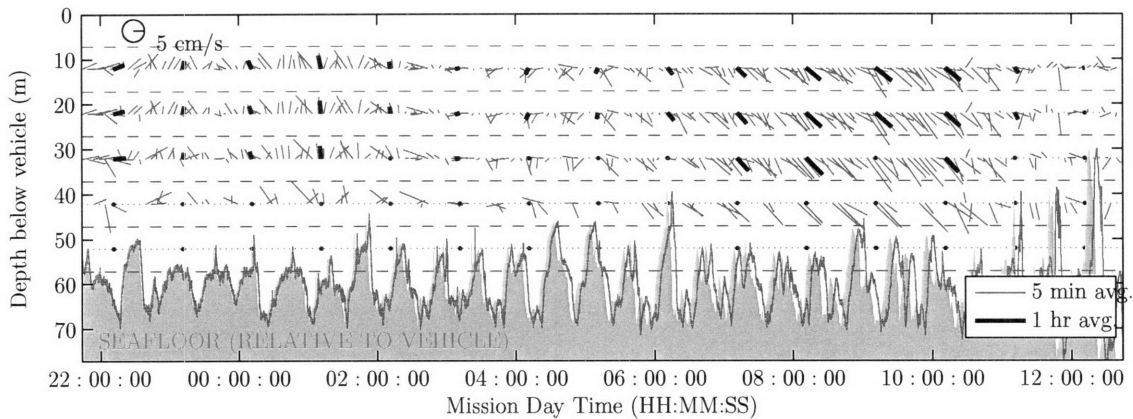


Figure 4-17: LBL-derived crossflow magnitude and direction profiles versus time, shown as 5 min and 1 h averages. The 5 min averages show considerably more variability than indicated by the same data processed using bottom-track velocity and shown in Fig. 4-15; however, the 1 h averages agree well.

at ABE's cruising speed of approximately 0.6 m/s. A 1 cm/s error is substantial when compared to typical observed crossflow speeds of 5 cm/s. Fortunately, comparisons of VOG computed from LBL to bottom-track VOG (when available) indicate that increasing the length of averaging intervals to span multiple tracklines on reciprocal headings substantially reduces error. This suggests a residual hard iron error [14]. Heading error due to hard iron offset will approximately cancel on reciprocal headings because hard-iron-induced heading errors are approximately sinusoidal with a period of 2π .

Figure 4-17 compares 5 min averages of LBL-derived crossflow velocity to 1 h averages for the same dive as shown in Fig. 4-15. The considerably increased temporal variability in the 5 min averages suggested by Fig. 4-17 over Fig. 4-15 indicates contamination from residual heading error. However, 1 h averages agree well with bottom-track derived crossflows on the same timescale; of course, a longer averaging interval engenders a loss of resolution as shown in Fig. 4-18. Crossflow measurements for dives lacking bottom-track are shown in App. F as 1 h averages.

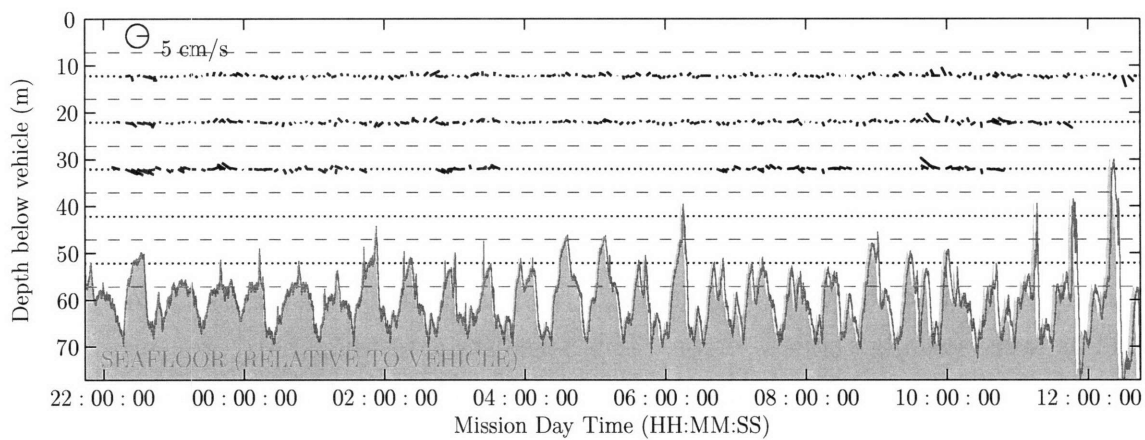


Figure 4-18: Crossflow error velocity computed by comparing interpolated 1 h average LBL-derived crossflow estimates to 5 min average bottom-track-derived crossflow estimates. Increasing averaging time from 5 min to 1 h improves the accuracy of LBL-derived crossflow estimates, but at the cost of reduced resolution.

4.5 Uncertainty in Seafloor Vent Location

This section develops the constraints on source location attained from individual detections of buoyant effluent considered in isolation. Intuitively, the spatial ambiguity associated with a single detection will depend on the width of the plume at the detection location, the strength of ambient currents (which advect the plume away from its source), as well as numerous secondary factors including source parameters, bathymetry, local hydrography, and the uncertainties associated with each. Because of the importance of horizontal advection in plume evolution, an ancillary goal of this section is to quantify the achievable reduction in source location uncertainty attained by measuring water current velocity on board the surveying vehicle. In principle, knowledge of current velocity should reduce the area of seafloor from which a plume could have emanated by constraining the source to lie upwind of the detection location. However, the achievable performance improvement will depend on the relative importance of advective versus dispersive processes in determining plume evolution and on the accuracy and availability of a plume model to predict that influence.

Presumably because of the much smaller range of space and time scales involved, BP detections have proven more amenable to interpretation in terms of source location than NBP detections, especially with respect to crossflows.³⁴ For this reason I restrict my attention in this section to the analysis of BP detections. Even for BP detections, the dependence on a model for describing BP rise is potentially problematic for two reasons: (1) source parameters cannot be known a priori; (2) it is impossible to predict turbulent plume tracer concentrations on the short timescales observed by a surveying AUV passing through a BP. Nevertheless, the modeling efforts presented in this section enable a key inference to be drawn from the anomaly maps of the previous section, namely the portion of seafloor that likely contains the source of each BP detection.

This section approaches vent localization from an inverse perspective. To apply the OG mapping methods of Part I to this problem, ultimately a forward model

³⁴Lavelle [67] has shown it is possible to invert a long record of current velocity and potential temperature in the NBP for the source strengths of hypothetical vent fields on a uniform grid; however, current records collected by the ABE AUV in our field operations have generally been too short to employ such a methodology, the cumulative advected distance rarely exceeding trackline spacing for phase-1 dives within the non-buoyant plume.

is required. Most of the aspects of the required model are developed here from an inverse perspective, with the remaining elements required to formulate the forward perspective left for the next chapter.

4.5.1 Buoyant Plume Interception

Figure 4-19 shows schematically an AUV intercepting a buoyant hydrothermal plume at some source relative vehicle height h_{vs} . The plume outline depicts rise due to buoyancy flux at the seafloor, plume spread due to self-generated and ambient turbulence and horizontal advection by ambient background currents. Because the volume of fluid ejected from the source is insignificant relative to the entrained fluid within most of the plume, the plume will acquire the horizontal momentum of the ambient fluids within a few source-diameters of rise height [20]. Detections of buoyant plume effluent are likely when the vehicle passes near the plume centerline; however, the information these detections carry about seafloor source location is obscured by plume spread. Even assuming perfect knowledge of the shape of the plume centerline, as vehicle height above the seafloor increases, the spatial extent of seafloor from which that plume could have emanated grows. However, the potential benefit attained by measuring ambient current velocity also increases because effluent will have advected farther from the source by the time it attains the height of the vehicle.

Independently of vehicle height, the physics governing plume rise affect the relative value of measuring currents. Interception of a faster rising plume will occur closer to the source, whereas advection will have longer to act on slower rising fluid, advecting it farther from the source before interception for the same vehicle height. Clearly the relative magnitude of these velocities is important. Recall from § 4.1 that plume rise is also affected by background stratification, source buoyancy flux, source chemistry, and source diameter [81, 82, 84, 110]. Of these, only the first can be known a priori.

Usually a surveying vehicle will only be able to measure a vertical profile of current velocity at its own location.³⁵ Assuming horizontal homogeneity, temporal variation in currents is easily accommodated. Over sufficiently large horizontal distances,

³⁵ADCPs carried on AUVs are typically oriented vertically (e.g. ABE and REMUS [1]) and can thus acquire vertical current profiles. ADCPs can be oriented horizontally, however, because the deep ocean is fairly homogeneous on short horizontal length-scales, the range would have to be significant. Furthermore, a horizontal orientation could result in bottom interference because of the 20° angular between the ADCP's beams.

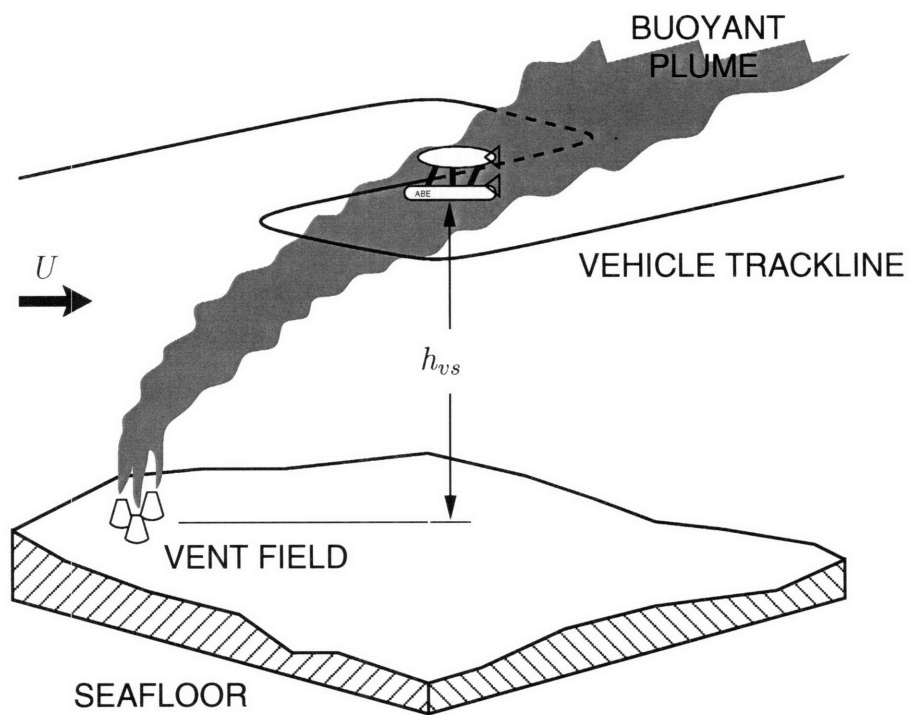


Figure 4-19: Buoyant hydrothermal plume interception by a vehicle surveying at a source-relative height h_{vs} .

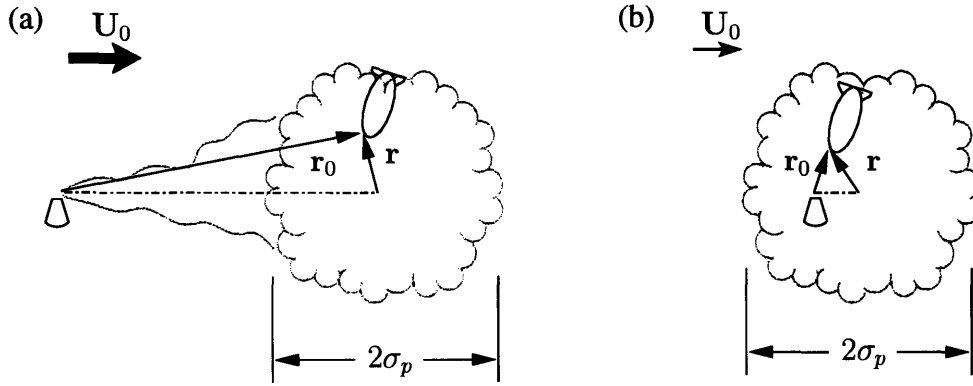


Figure 4-20: The effect of ambient crossflow strength U_0 on the location of a plume at vehicle depth assuming constant rise rate W_0 : **(a)** strong crossflow ($U_0/W_0 > 1$); **(b)** weak crossflow ($U_0/W_0 < 1$). Higher intensity crossflows advect the plume over larger horizontal distances for the same vertical source-vehicle separation h_{vs} . Within the plume, the average distance to the plume centerline $E[|\mathbf{r}|]$ is less than the average distance to the plume source on the seafloor $E[|\mathbf{r}_0|]$; however, wide plumes and weak advection reduce the magnitude of the difference.

bathymetry and other influences will degrade the quality of this assumption. Furthermore, without assuming some kind of model for the spatio-temporal variability of the currents, it is impossible to separate spatial variations from temporal variations, the two being coupled through the motion of the vehicle.³⁶

4.5.2 Distribution of BP Detection Location

Figure 4-20 schematically illustrates the effect of ambient crossflow strength $|\mathbf{U}_0|$ on the source-relative location of a buoyant plume. Effluent is assumed to issue from a point source and rise at a constant rate W_0 thereafter.³⁷ Letting σ_p denote a characteristic plume radius, case (a) corresponds to dominant advection ($\sigma_p < U_0 h_{vs}/W_0$), and case (b) corresponds to weak advection with plume evolution dominated by plume spread ($\sigma_p > U_0 h_{vs}/W_0$). Higher intensity crossflows advect the plume over larger horizontal distances for the same vertical source-vehicle separation h_{vs} .

³⁶For processes with known spatio-temporal spectra, Zhang et al. [137] developed the concept of a “mingled-spectrum” to describe the mapping of spatial and temporal variation onto a single temporal axis through the vehicle’s velocity.

³⁷This assumption is useful here to establish intuition on the basis of the simple geometric relationships that result. A justification is offered subsequently in § 4.5.3

Figure 4-20 also defines two scalar quantities r_0 , the horizontal distance between a vent and vehicle location upon detection of the plume emanating from that vent, and r , the horizontal distance between the plume's centerline location at vehicle depth and vehicle location upon detection. These distances are the key to assessing the value of measuring U_0 in situ. Because the plume evolution itself is a random process, these quantities are both inherently random. In principle, knowledge of centerline evolution and crossflow should enable the uncertainty associated with the source location to reflect the statistics of r rather than r_0 . Thus the attainable reduction in source location uncertainty only becomes significant when crossflow intensity dominates plume growth, as in case (a) of Fig. 4-20.

For all crossflow intensities and plume widths, the triangle inequalities applied to r_0 , r , and centerline length $U_0 h_{vs}/W_0$ yield the constraints:

$$\begin{aligned} r_0 + r &\geq \frac{U_0 h_{vs}}{W_0} \\ |r_0 - r| &\leq \frac{U_0 h_{vs}}{W_0} . \end{aligned} \tag{4.12}$$

These bounds are purely geometric, and apply subject to the assumption of a constant rise rate W_0 so long as crossflows remain invariant. These bounds reveal themselves in Figure 4-21 which shows realizations of r vs r_0 attained for 20 simulated detections, using (a) a relatively strong non-dimensional crossflow U_0/W_0 and (b) a relatively weak non-dimensional crossflow. The samples were generated under the assumption that r is Rayleigh distributed about the centerline location. While on average $E[r] \leq E[r_0]$, for an individual detection the associated r and r_0 may be such that $r_0 < r$. Of course there is no way of knowing that the source is in fact closer than the plume centerline without a priori knowledge of source location.

4.5.3 Dependence on source-relative vehicle height h_{vs}

Under favorable crossflow conditions, realizing a reduction in source location ambiguity from measurement of the crossflow still requires accurate prediction of plume centerline location at vehicle depth. That prediction will be subject to modeling error as well as to error in estimates of the crossflow due to either direct measurement error, or the spatial variability thereof. Furthermore, both the attainable reduction

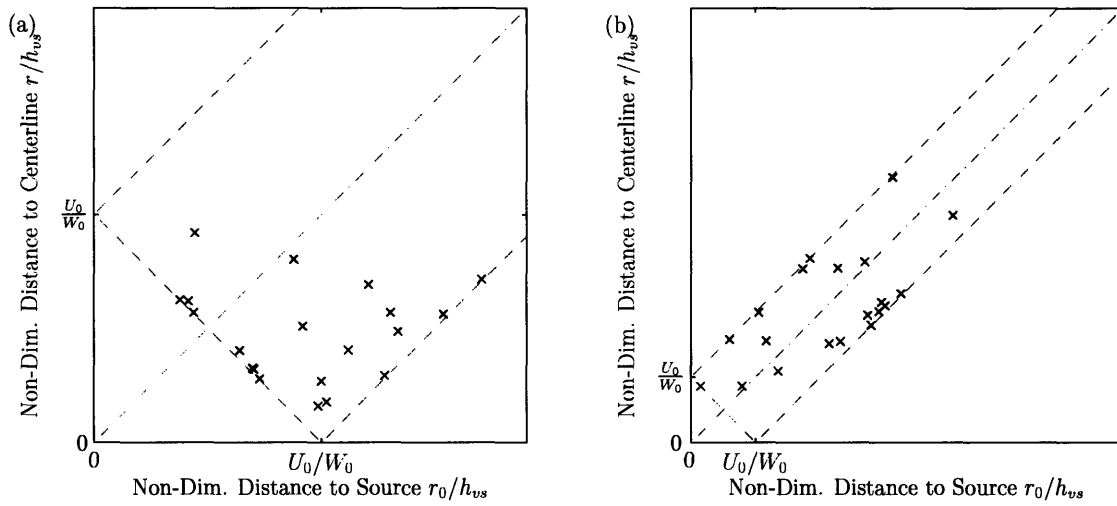


Figure 4-21: Non-dimensional distance to plume centerline versus non-dimensional distance to source for 20 simulated detection locations assuming Rayleigh distribution about plume centerline: **(a)** strong crossflow ($U_0/W_0 > 1$); **(b)** weak crossflow ($U_0/W_0 < 1$). The bounds specified by (4.12) are indicated by dashed lines. As crossflow intensity decreases, these bounds move toward the 1 : 1 line, reducing any benefit attained from accurate prediction of centerline location. Similarly, as plume width increases the sample points spread out along the r axis, filling a larger portion of the band around the 1 : 1 line, thereby effectively reducing the distinction between r and r_0 as well.

in source location ambiguity as well as the fidelity of any centerline prediction should be functions of source-relative vehicle height h_{vs} : a plume encountered at larger h_{vs} will have had more time to advect farther from its source in the same crossflow. This intuitive idea is supported by Fig. 4-22 which shows the distances at which BP detections were observed versus vehicle height. Finally, effective source size or vent field extent will bound the achievable resolution. I turn my attention to these factors next. The ensuing development also affords an opportunity to assess the assumptions of a constant rise rate and of a Rayleigh distribution for r employed in Figs. 4-20 & 4-21.

I now develop a model for the expected distribution of source-relative detection locations from the elementary perspective of a discrete particle of vent fluid ejected from a hydrothermal source. From § 4.1, the factors affecting the trajectory of the particle are (1) buoyancy of the particle and the surrounding fluid; (2) turbulence, mostly self-generated by the plume's own buoyancy; (3) advection by horizontal crossflow. Contributions from both buoyancy and turbulence will determine the vertical motions undertaken by the particle. Considerable simplification is attained, however, by assuming that buoyancy dominates and furthermore that it acts to deterministically increase particle height above the source as a function of time. Finally, I assume that crossflows (composed primarily of diurnal tidal currents) can be regarded as constant over the 1 h timescale of buoyant plume rise. Under these assumptions, the particle's position \mathbf{x}_i as a function of time t and time-since-emission τ_e is:

$$\mathbf{x}_i(t, \tau_e) = \mathbf{x}_s + \mathbf{U}(t) \cdot \tau_e + h_{CL}(\tau_e) \hat{\mathbf{k}} + \int_{t-\tau_e}^t \mathbf{u}_i(t - \tau) d\tau, \quad (4.13)$$

where \mathbf{x}_s denotes the particle's origin from a source on the seafloor, $\mathbf{U}(t)$ denotes horizontal advective currents assumed constant over the interval $[t - \tau_e, t]$ and henceforth denoted \mathbf{U}_0 , $\hat{\mathbf{k}}$ denotes the unit vertical (positive down), $h_{CL}(\tau_e)$ denotes centerline rise height as a function of time-since-emission, and $\mathbf{u}_i(t)$ denotes the turbulent motions executed by the particle. These are limited to horizontal motions about plume centerline by assumption, i.e. $\mathbf{u}_i^\top \hat{\mathbf{k}} = 0$. Two elements of (4.13) are subject to randomness: \mathbf{x}_s , because vent fields have finite size, and $\mathbf{u}_i(t)$, because of the stochastic nature of turbulence.

While an individual hydrothermal orifice might be reasonably approximated by a

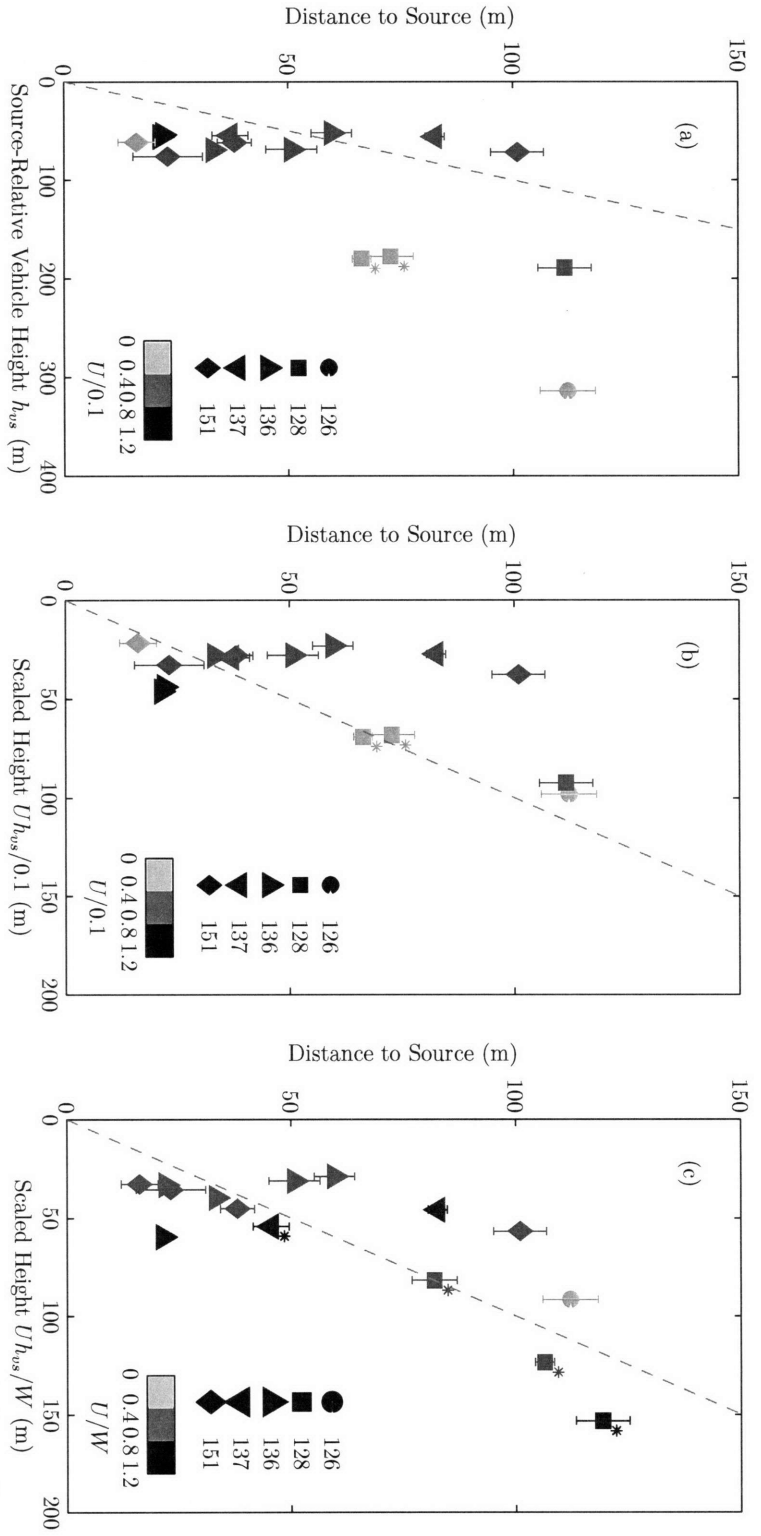


Figure 4-22: Measured distance to source r_0 versus (a) source relative vehicle height h_{vs} for all studied ABE dives; (b) vehicle height scaled by measured crossflow intensity and canonical rise rate, $U_0 h_{vs} / W_0$. Measured crossflow intensity is indicated by the color of each mark. Detections from five dives are plotted, indicated by the shape of the mark used. All plots indicate an increase in r_0 with vehicle height, although considerable scatter is also evident at vehicle heights of ~ 50 m where BP detections were encountered most often. Assuming a constant rise rate, scaling of h_{vs} by crossflow intensity over rise rate should account for variations in crossflow intensity thereby making most detections fall nearer to the 1 : 1 line indicated in each plot by a dashed line. Using the average measured rise rate as in (c) as opposed to the canonical value as in (b) appears to improve the trend. In case of multiple nearby sources, detections were assigned to sources based on proximity to a circle of radius $U_0 h_{vs} / W_0$ using $W_0 = 0.1$ m/s (panels a&b) or measured rise rate (panel c). Starred data points indicate those detections for which another source was closer to the detection location than the assigned source. Error bars indicate a weighted measure of the width of each detection.

point source, nearby orifices on a single hydrothermal structure, or multiple venting structures in a field may produce plumes that coalesce as they rise. To account for this, let the horizontal component of particle emission location ${}^h\mathbf{x}_s$ follow a normal distribution:

$${}^h\mathbf{x}_s \sim \mathcal{N} \{ {}^h\bar{\mathbf{x}}_s, \sigma_s^2 \mathbf{I}_{2 \times 2} \} , \quad (4.14)$$

where ${}^h\bar{\mathbf{x}}_s$ denotes the center of the vent field, and σ_s a measure of field extent. I assume the vertical emission location z_s is known approximately from bathymetry and constrained to a small depth interval (relative to σ_s) such that it can be regarded as deterministic. The horizontal field extent represents a lower bound on the achievable precision of source locations inferred from mid-water column detections.

It remains to describe the distribution of the horizontal components of the turbulent motions ${}^h\mathbf{u}_i(t)$. This can be done by linking the random motions of individual particles to the known time-averaged concentration profile across a plume. The probability density function of an individual particle's location $p(\mathbf{x}_i; \tau_e)$ is linked to ensemble average concentration $\chi(\mathbf{x}, \tau_e)$ through the relationship [20]:

$$\chi(\mathbf{x}, \tau_e) = Q p(\mathbf{x}; \tau_e) d\mathbf{x} \quad (4.15)$$

where $p(\mathbf{x}; \tau_e)$ denotes the particle position probability density evaluated at \mathbf{x} , and Q denotes the total mass of particles released. The probability densities associated with each particle must be identical functions of τ_e for (4.15) to apply. The analogous form of (4.15) for a steady, continuous release of particles is, by superposition of the ensemble average fields [20],

$$\chi(\mathbf{x}) = q \int_0^\infty p(\mathbf{x}; \tau) d\tau \cdot d\mathbf{x} , \quad (4.16)$$

where the ensemble average no longer contains any time dependence and is equivalently a time-average concentration.³⁸ Total particle mass Q has been replaced by a steady mass release rate q to accommodate this continuous release perspective, as opposed to the release of an discrete number of particles.³⁹

³⁸Ensemble average connection to time-average for continuous release.

³⁹I assume a steady release rate throughout this work. Hydrothermal venting is known to vary in intensity, though the extent of this variation is poorly constrained. Other types of geophysical flows like cold seeps may be subject to considerable variation due to tidally induced pressure changes.

Assuming independence of the three components of particle position,

$$\chi({}^h\mathbf{x}, z) = q \int_0^\infty p({}^h\mathbf{x}; \tau) p(z; \tau) d\tau \cdot d\mathbf{x} . \quad (4.17)$$

By assumption, vertical particle position z_i is a deterministic function of time-since-release τ_e so that

$$p(z; \tau_e) = \delta(z - h_{\text{CL}}(\tau_e)) . \quad (4.18)$$

Further assuming that $h_{\text{CL}}(\tau_e)$ is invertible, the integrand of (4.17) will be zero everywhere except at $\tau_e = h_{\text{CL}}^{-1}(z_s - z)$ yielding

$$\chi({}^h\mathbf{x}, z) = q p({}^h\mathbf{x}; h_{\text{CL}}^{-1}(z_s - z)) d\mathbf{x} . \quad (4.19)$$

Turbulent motions are three-dimensional, thus the assumption of a deterministic z_i is a considerable simplification. However, their effects on time-averaged concentration $\chi(\mathbf{x})$ are minor if the ascending plume is narrow such that the concentration gradient in the z -direction is shallow relative to the horizontal gradient. In that case, turbulent vertical motions will tend to cancel one another out in a frame moving with the bulk flow, as effluent from above will be brought down at roughly the same rate as effluent being brought up from below.⁴⁰

Observations of laboratory plumes indicate that time-averaged concentration profiles across buoyant plumes are approximately Gaussian [82], therefore it follows from (4.19) that ${}^h\mathbf{x}_i(z)$ will be normally distributed. Thus only the covariance of the integral containing \mathbf{u}_i in the particle position model (4.13) is required to completely specify the distribution of $\mathbf{x}_i(\tau_e)$. In general, this covariance will be a function of time-since-release τ_e . Integration of the turbulent velocities ${}^h\mathbf{u}_i$ will tend to de-correlate the components of the resulting vector such that a τ_e -dependent covariance describing

Applying the approaches developed in this thesis to systems with significant temporal variability in source characteristics would require models capable of predicting this variation, a daunting task, or else models that do not attempt to predict the aspects of plumes affected by varying source fluxes. Sources that move or turn off completely imply a time-dependent true map and cannot be accommodated without augmenting the state space of the map.

⁴⁰This argument is usually employed to justify the so-called “slender-plume” approximation to turbulent diffusion of a passive tracer. The argument requires explicit integration of (4.15) using an assumed (Gaussian) form for $p(\mathbf{x}; \tau_e)$ and assuming the associated variance is known as a function of τ_e (e.g. [20]), but avoids the need for an a priori assumption of negligible axial diffusion.

plume width can be expected to have the form

$$\sigma_p(\tau_e) \mathbf{I}_{2 \times 2} = \text{cov} \left(\int_0^{\tau_e} {}^h \mathbf{u}_i(\tau) d\tau \right) , \quad (4.20)$$

where the dependence on time t has been dropped since the statistics of ${}^h \mathbf{u}_i$ are a function of τ_e alone. Below, I specify $\sigma_p(\tau_e)$ directly based on empirical observations of buoyant hydrothermal plumes, obviating any need to specify the statistics of \mathbf{u}_i , in effect relying on the Central Limit Theorem to ensure that the resulting distribution of particle position will be Gaussian regardless of the details of \mathbf{u}_i .⁴¹ On the basis of this argument, horizontal particle position will be distributed according to

$${}^h \mathbf{x}_i(\tau_e) \sim \mathcal{N} \left\{ {}^h \bar{\mathbf{x}}_s + {}^h \mathbf{U}_0 \cdot \tau_e, \left(\sigma_p^2(\tau_e) + \sigma_s^2 \right) \mathbf{I}_{2 \times 2} \right\} . \quad (4.21)$$

Finally, the dependence on τ_e can be replaced by a dependence on z via the transformation $\tau_e = h_{\text{CL}}^{-1}(z_s - z)$ (assumed to exist over the range of interest, $z < z_s$):

$${}^h \mathbf{x}_i(z) \sim \mathcal{N} \left\{ {}^h \bar{\mathbf{x}}_s + {}^h \mathbf{U}_0 \cdot h_{\text{CL}}^{-1}(z_s - z), \left(\sigma_p^2(h_{\text{CL}}^{-1}(z_s - z)) + \sigma_s^2 \right) \mathbf{I}_{2 \times 2} \right\} . \quad (4.22)$$

Up to this point, no specific reference has been made to any models of buoyant plume rise, only to the general characteristics thereof. To proceed, the functions $h_{\text{CL}}(\tau_e)$ and $\sigma_p(\tau_e)$ need to be specified. A crude, but nonetheless illuminating perspective is afforded by assuming a constant rise rate W_0 , and furthermore that plume radius grow linearly with height h_{CL} :

$$h_{\text{CL}}(\tau_e) = W_0 \tau_e \quad (4.23a)$$

$$\sigma_p(h_{\text{CL}}) = \frac{1}{2} a h_{\text{CL}} , \quad (4.23b)$$

where a represents a constant of proportionality describing the increase in plume radius per unit rise.⁴² Though crude, these expressions can be justified to a degree

⁴¹The relationship between $\sigma_p \tau_e$ and the covariance of ${}^h \mathbf{u}_i$ is easily attained under the simplifying assumption that ${}^h \mathbf{u}_i$ is a (non-stationary) continuous-time white noise process. Straightforward manipulations lead to $\text{cov}(\mathbf{u}_i) = \frac{d}{d\tau_e} \sigma_p^2(\tau_e) \mathbf{I}_{2 \times 2}$. For example, if $\text{cov}(\mathbf{u}_i)$ were constant (as it would be for molecular diffusion), then $\sigma_p(\tau_e) \propto \sqrt{\tau_e}$, yielding the familiar (e.g. [20]) \sqrt{t} law for the growth in width of a passive plume in the far field of a continuously releasing source.

⁴²The leading factor of $\frac{1}{2}$ reflects the scaling between plume radius for a top hat cross-section

with respect to the classical “top-hat” model of time-averaged buoyant plume rise originally formulated by Morton, Taylor, and Turner [89], and applied subsequently to hydrothermal plumes by Speer and Rona [110] and others.

The essence of the MTT model is an artificial distinction between the ambient environment and plume fluid [20]. That distinction enables the closure of integral conservation equations for the mass, momentum and buoyancy of the rising plume via the assumption that such plumes grow by incorporating ambient fluid into themselves as they rise. The process by which this occurs is known as turbulent entrainment. Because the entrainment results from the plumes own motion, a reasonable assumption is that it should occur at a rate proportional to its local vertical velocity.⁴³ The associated constant of proportionality, known as the entrainment coefficient and usually denoted α , has a typical value of $\alpha = 0.07$ (e.g. [82]).

In unstratified surroundings, the MTT model predicts indefinite rise and linear growth in plume radius with respect to height. In uniformly stratified surroundings (constant Brunt-Väisälä buoyancy frequency) buoyant plumes eventually achieve a terminal rise height, though their radii still grow approximately linearly up to roughly 90% of the level of neutral buoyancy before beginning a more rapid expansion [119]. On this basis, the assumption of linear growth in plume radius (4.23a) is a reasonable approximation, especially for plumes encountered beneath the NBP. McDuff [83] reports a radial growth rate of 10 cm per meter of rise for typical plumes emanating from the hydrothermal vents of the Main Endeavour Field located in the Northeast Pacific giving $a = 0.1$ m/m.

The assumption (4.23b) of a constant rise rate is rather more difficult to defend as it ignores the dynamic balance between plume growth through entrainment, vertical momentum, and buoyancy forces acting on the ascending plume. Modeling studies carried out by Speer and Rona [110] for canonical Pacific and Atlantic plumes predict a nearly instantaneous initial vertical acceleration followed by rapid decay over the first few tens of meters of rise, then relatively slow decay throughout most of a plume’s rise, until a final period of rapid decay beyond the level of neutral buoyancy. They give a figure of 10 cm/s for both Atlantic and Pacific plumes as describing vertical

versus standard deviation for a Gaussian plume cross-section [47]

⁴³Turner [119] presents an overview of the many applications of the entrainment assumption to geophysical flows as well as a discussion of the mechanism itself.

velocity throughout most of the vertical extent of the plume. I will refer to this figure, $W_0 = 0.1$ m/s, subsequently as the “canonical rise rate” for buoyant hydrothermal plumes. Thus the assumptions (4.23), though rough, are not unreasonable on the basis of theoretical predictions, the first of which applies to all buoyant plumes, and the latter specifically to hydrothermal plumes.⁴⁴ MTT-based modeling efforts specific to hydrothermal plumes in the buoyant phase (e.g. [81,84,107,110]) should in principle produce more accurate results than the simplistic perspective taken above; however, these are of dubious utility in the present context where source parameters cannot be known a priori.

Returning now to the probability distribution of the location of a particle emitted from a hydrothermal source, the assumptions (4.23) provide a functional form in terms of source-relative vehicle height $h_{vs} = z_s - z_v$ at the hypothetical interception of particle i :

$${}^h\mathbf{x}_i(h_{vs}) \sim \mathcal{N} \left\{ {}^h\bar{\mathbf{x}}_s + {}^h\mathbf{U}_0 \frac{h_{vs}}{W_0}, \left(\frac{1}{4} a^2 h_{vs} + \sigma_s^2 \right) \mathbf{I}_{2 \times 2} \right\} . \quad (4.24)$$

Intuitively, a parallel exists between the uncertainty associated with an emitted particle’s location given its origin and the uncertainty associated with a particle’s origin given its location in the water column. Since centerline evolution and crossflow play important roles in determining particle location, knowledge of these processes should enable a reduction in the uncertainty associated with source location; however, neither can be known exactly. A prediction of particle location $\hat{\mathbf{x}}_i$, or equivalently an estimate of source location $\hat{\mathbf{x}}_s$ will in either case be subject to crossflow measurement error and centerline model error in addition to the inherent uncertainty introduced by plume growth and source extent indicated by (4.24). I will restrict my attention to errors in the measurement of crossflow \mathbf{U}_0 , since its effect is of principal importance

⁴⁴In light of the entrainment assumption, the consistency of the assumptions (4.23) might also be of concern. According to the MTT model, conservation of mass expressed for a horizontal disk of plume with area $A(z)$ and vertical velocity $W(z)$ gives $d/dz(AW) = 2\sqrt{\pi}\alpha A^{1/2}W$ where α denotes the entrainment coefficient. Assuming linear growth in plume radius parameterized by a leads to a solution for plume rise rate $W(z) = \text{const.} \cdot z^{\frac{-2}{a-2\alpha}}$ (not the usual MTT solution of course, which requires simultaneous solution of equations for conservation of mass, momentum, and buoyancy). On the basis of this equation, if $a = 2\alpha$, rise rate would in fact be constant even under the assumption of a linear growth in plume radius. The values for a and α quoted in the text (0.1 and 0.07, respectively) are in rough agreement with this condition.

in determining the value of measuring currents aboard an AUV. Let $\hat{\mathbf{U}}_0$ denote the crossflow estimate, and $\tilde{\mathbf{U}}_0 \triangleq \hat{\mathbf{U}}_0 - \mathbf{U}_0$ denote the associated error:

$$\begin{aligned}\tilde{\mathbf{U}}_0 &\sim \mathcal{N}\left\{\mathbf{0}, \sigma_{\tilde{U}_0}^2 \mathbf{I}_{2 \times 2}\right\} \quad (\text{measured}) \\ \tilde{\mathbf{U}}_0 &\sim \mathcal{N}\left\{\mathbf{0}, \sigma_{U_0}^2 \mathbf{I}_{2 \times 2}\right\} \quad (\text{not measured})\end{aligned}\tag{4.25}$$

where the covariance associated with not measuring currents $\sigma_{U_0}^2 \mathbf{I}_{2 \times 2}$ reflects the covariance of the crossflow itself, while $\sigma_{\tilde{U}_0}^2 \mathbf{I}_{2 \times 2}$ reflects measurement error, and presumably $\sigma_{U_0} > \sigma_{\tilde{U}_0}$.

An unbiased estimate of horizontal particle location at depth $z_i = z_s + h_{\text{CL}}$ for known source location $\bar{\mathbf{x}}_s$ is:

$${}^h\hat{\mathbf{x}}_i = {}^h\bar{\mathbf{x}}_s + \hat{\mathbf{U}}_0 \frac{h_{\text{CL}}}{W_0} .\tag{4.26}$$

Assuming the sources of uncertainty described above are uncorrelated, the error in predicted particle location $\tilde{\mathbf{x}}_i \triangleq \hat{\mathbf{x}}_i - \mathbf{x}_i$ will be distributed according to

$$\begin{aligned}{}^h\tilde{\mathbf{x}}_i &\sim \mathcal{N}\left\{\mathbf{0}, \left(\left(\frac{\sigma_U^2}{W_0^2} + \frac{a^2}{4}\right) h_{\text{CL}}^2 + \sigma_s^2\right) \mathbf{I}_{2 \times 2}\right\} \\ \tilde{z}_i &= 0 \quad (\text{by assumption})\end{aligned}\tag{4.27}$$

where $\sigma_U = \sigma_{U_0}$ without crossflow measurements, or else $\sigma_U = \sigma_{\tilde{U}_0}$. The covariance in (4.27) is functionally dependent on two constant parameters: σ_s and $\beta \triangleq \left(\frac{\sigma_U^2}{W_0^2} + \frac{a^2}{4}\right)$ which consists of the lumped contribution from imperfect knowledge of advection plus plume growth. At low altitudes, source size σ_s dominates the uncertainty, whereas as height increases, an approximately linear dependence on h_{CL} develops, scaled by β .

Figure 4-23 shows distance to source r_0 and distance to predicted centerline location r computed for the BP interceptions plotted in Fig. 4-22. According to particle position model (4.27), these data will follow a Rayleigh distribution with h_{vs} -dependent parameter $(\beta h_{vs}^2 + \sigma_s^2)^{\frac{1}{2}}$, and different β according to:

$$\begin{aligned}r_0 : \quad \beta &= \frac{\sigma_{U_0}^2}{W_0^2} + \frac{a^2}{4} \\ r : \quad \beta &= \frac{\sigma_{\tilde{U}_0}^2}{W_0^2} + \frac{a^2}{4} .\end{aligned}\tag{4.28}$$

Table 4.1: Maximum likelihood estimates (95% confidence) of lumped plume spread β and characteristic vent field size σ_s . The value for σ_s with $\beta \equiv 0$ represents a baseline computed by assuming no dependence of uncertainty on vehicle height. Because of the small number of BP detections at heights above ~ 50 m, this estimate is likely biased toward a smaller value.

Distance	Source Assignment	β	σ_s (m)
r_0	$\text{argmin}_s \bar{\mathbf{x}}_s - \mathbf{x}_v $	$\equiv 0$	43 [34, 57]
r_0	$\text{argmin}_s \bar{\mathbf{x}}_s - \mathbf{x}_v - U_0 h_{vs}/0.1 $	0.11 [-0.04, 0.26]	30 [13, 48]
r	$\text{argmin}_s \bar{\mathbf{x}}_s - \mathbf{x}_v - U_0 h_{vs}/0.1 $	0.000 [-0.011, 0.012]	26 [19, 33]

These data were used to compute maximum likelihood estimates of the parameters β and σ_s ; results are given in Table 4.1. The 95% confidence intervals for β are wide reflecting the sparsity of BP detections for $h_{vs} \gtrsim 50$ m. Nevertheless, as expected, the β predicted for r_0 with crossflow treated as random is larger than for r . The estimated values of σ_s , which are theoretically identical in either case, agree well.

Especially considering the wide confidence intervals, a comparison to order-of-magnitude estimates of σ_s and especially β is instructive. Field extent σ_s is obviously site-specific; however, values computed for σ_s in Table 4.1 probably reflect both the crudeness of the model (4.27) as well as variability between the five different vent fields discovered over the course of the dives considered. The parameter β is more easily predicted. A $\sigma_{U_0} = 0.04$ m/s is consistent with a Rayleigh distribution of crossflow intensity such that $E[|U_0|] = 5$ cm/s. A $\sigma_{\tilde{U}_0} = 0.01$ m/s overestimates the attainable precision of ADCP derived current measurements, but I believe represents a reasonable estimate of the influence of spatial variability in crossflow on predicted plume location. Along with $a = 0.1$ from above, these numbers result in values of 0.16 and 0.013 for β associated with r_0 and r respectively, in broad agreement with the maximum likelihood estimates in Table 4.1.

Performance Metric

The preceding discussion derived distributions for distance-to-source r_0 and distance-to-plume-centerline r from detection location as functions of vehicle height h_{vs} and employed data from ABE dives to determine values for the unknown parameters.

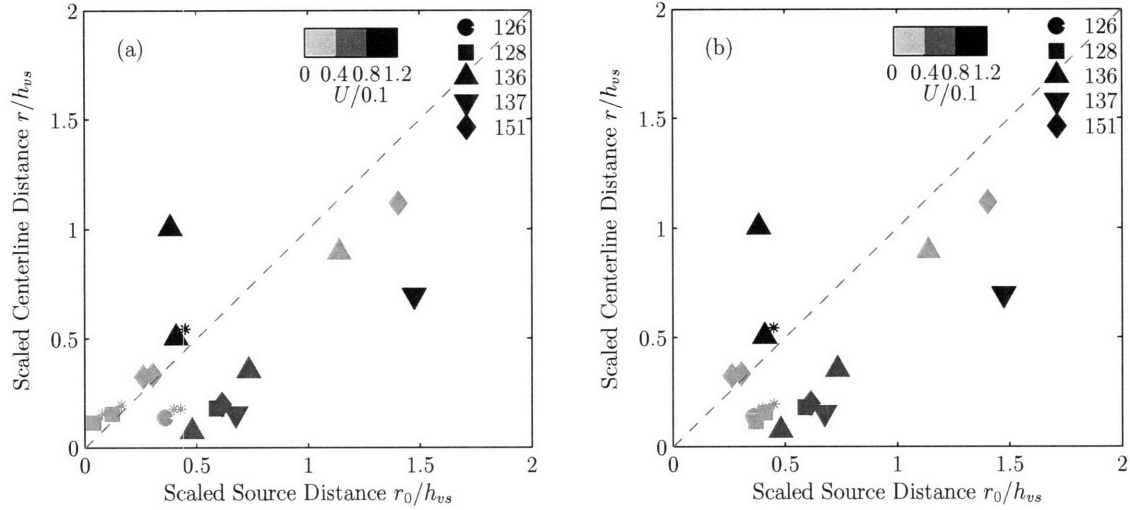


Figure 4-23: Non-dimensional distance to plume centerline r/h_{vs} versus non-dimensional distance to source r_0/h_{vs} for BP detections from five ABE dives (indicated by the mark shape): **(a)** source assignment for r_0 according to minimum distance; **(b)** by proximity to a circle of radius $U_0 h_{vs}/0.1$ around the detection location. Centerline locations were predicted by numerically integrating measured crossflow. Average measured crossflow intensity is indicated by the color of each mark. In both panels, most points occur either near or below the 1 : 1 line indicating that, on average, a reduction in source location ambiguity is attainable by predicting centerline location at vehicle height. Distances measured during periods of low crossflow intensity (lightest gray) show only minimal improvement as expected since plumes should rise nearly vertically under these conditions. Medium crossflow intensities (dark gray) produced the best and most consistent improvement. Distances measured during periods of high intensity crossflow (black) produced scattered results, even though the best improvements should have been observed during these times. This may reflect the inadequacy of the constant rise-rate assumption to account for the interplay between crossflow and rise rate [84]. Starred data points indicate those detections for which the source assignment used to compute r and r_0 was in disagreement. The crossflow record for the one double-starred point from ABE-136 was insufficiently long to predict centerline location by integration of measured crossflow, and a centered average was used instead.

This section derives a non-dimensional metric for the reduction of source location ambiguity attained by using crossflow measurements to help constrain source location.

Let each detection be associated with a patch of seafloor from which it could have emanated of some characteristic area. Assuming known $h_{\text{CL}} = h_{vs}$,⁴⁵ the uncertainty associated with source location given the location of a buoyant plume detection is simply the uncertainty associated with detection location given source location developed above. Thus accurate prediction of centerline trajectory, enabled by on-board measurement of crossflow, should reduce the characteristic seafloor area from $\sigma_{r_0}^2 \triangleq \text{var}(r_0)$ to the (smaller) characteristic area $\sigma_r^2 \triangleq \text{var}(r)$.

A normalized quantity describing the reduction in locational ambiguity attained by including an estimate of plume centerline in interpreting a water-column detection is

$$\eta_r \triangleq \frac{\sigma_{r_0} - \sigma_r}{\sigma_{r_0}} . \quad (4.29)$$

The normalized improvement $\eta_r \in [0, 1]$ represents the normalized reduction in the characteristic radius of the patch of seafloor expected to contain the source of the detected plume. From (4.27) & (4.28), η_r may be parameterized as:

$$\eta_r = \eta_r(\sigma_{U_0}, \sigma_{\tilde{U}_0}, a, W_0, h_{vs}, \sigma_s) . \quad (4.30)$$

Supposing crossflow is measured accurately such that dependence on $\sigma_{\tilde{U}_0}$ is eliminated⁴⁶, and setting $\sigma_{U_0} \propto U_0$ yields a simplified parameterization in terms of non-dimensionalized quantities:

$$\eta_r = \eta_r\left(\frac{U_0}{W_0}, a, \frac{h_{vs}}{\sigma_s}\right) . \quad (4.31)$$

⁴⁵Source-relative vehicle height h_{vs} , or equivalently centerline height at interception h_{CL} , will be known approximately from bathymetry. Considerable relief on length scales comparable to the uncertainty in source location will alter the shape of the patch of seafloor from which the detected plume emanated. This effect is automatically accommodated by the Occupancy Grid based stochastic mapping method applied to BP detections in Ch. 5.

⁴⁶From (4.28) measured accurately enough to be ignored requires $\frac{\sigma_{\tilde{U}_0}}{W_0} \ll \frac{a}{4}$.

The functional form of (4.31) is then

$$\eta_r = 1 - \left(\frac{\frac{a^2}{4} \left(\frac{h_{vs}}{\sigma_s} \right)^2 + 1}{\left(\frac{2}{\pi} \left(\frac{U_0}{W_0} \right)^2 + \frac{a^2}{4} \right) \left(\left(\frac{h_{vs}}{\sigma_s} \right)^2 + 1 \right)} \right)^{\frac{1}{2}}, \quad (4.32)$$

where the factor of $\frac{2}{\pi}$ arises from assuming a Rayleigh distribution for crossflow magnitude $|\mathbf{U}_0|$.⁴⁷

Eq. (4.32) is plotted in Fig. 4-24 for fixed $a = 0.1$ m/m. Attainable performance improves as crossflow speed increases and for increasing vehicle height, whereas larger characteristic vent field size σ_s reduces η_r . For the parameter estimates in Table 4.1 and at a typical observed crossflow velocity of 5 cm/s Fig. 4-24 indicates $\eta_r = 0.2$ for Phase-2 dives at 50 m height, and $\eta_r = 0.7$ for a maximum Phase-1 vehicle height of 300 m.

4.6 Conclusions

This chapter contains two principal contributions:

1. An automated procedure for the classification of hydrographic data as either background water unaffected by hydrothermal input, water contaminated by a non-buoyant plume (NBP), or buoyant plume (BP) water.
2. A parameterized model for the probability density function of the seafloor location of the source of a BP detected at some height above the seafloor.

The first of these was applied to a set of four measurements collected by the ABE vehicle to produce maps of plume activity suitable for autonomous interpretation by virtue of having segmented plumes from background. The method is grounded in the physical characteristics of hydrothermal plumes rather than in the specifics of each measurement and should be applicable to scalar measurements from sensors other than those studied here.

⁴⁷With $U_0 = E[|\mathbf{U}_0|]$, a Rayleigh distribution for $|\mathbf{U}_0|$ implies $\sigma_{U_0} = \sqrt{\frac{2}{\pi}} U_0$.

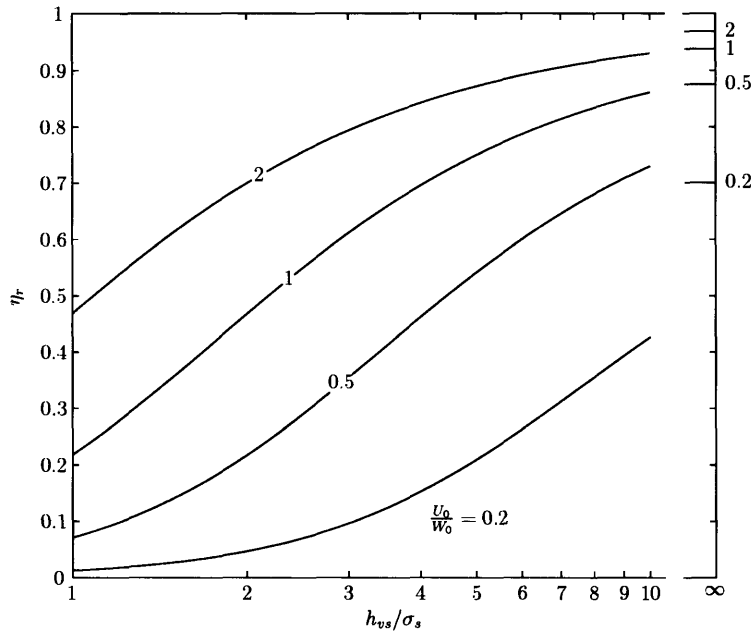


Figure 4-24: Attainable reduction in normalized source location ambiguity η_r from on-board measurement of crossflow as a function of non-dimensionalized crossflow speed U_0/W_0 and non-dimensionalized vehicle height h_{vs}/σ_s for fixed plume growth parameter $a = 0.1$. Attainable performance improves as crossflow speed increases and for increasing vehicle height. Larger characteristic vent field size σ_s reduces the attainable performance improvement. Marks on right side of the plot indicate performance improvement in the limit of a point source ($\sigma_s \rightarrow 0$), in which case η_r becomes independent of vehicle height.

Based on the estimated parameters of the model for uncertainty in source location, source locations can be estimated with respect to the locations of individual BP detections with a standard deviation of ~ 25 m at a vehicle height of 50 m (nominal Phase-2), increasing to ~ 40 m at a maximum Phase-1 height of 100 m, all assuming the availability of crossflow measurements from an on board ADCP. Without an ADCP for measuring crossflow, these numbers worsen to ~ 30 m and ~ 100 m, respectively.

Now consider these contributions in light of the questions posed at the end of § 4.2. Recall that these pertained to the *reliability*, *efficiency*, and potential for *autonomy* in the three-stage nested survey plan introduced in § 4.2.

Reliability: The first contribution listed above provides for detection of BP contact with a specified probability of false alarm. Based on the complete results provided for all dives in App. F, BP detections were reliable indicators of venting below, consistent with a low probability of false alarm. That is, active venting was discovered in all areas near BP detections (when that area was surveyed during a Phase-3 near-bottom dive). One notable exception was the southern-most BP detection during ABE-128. Further evidence of BP contact in this region was observed during the Phase-2 dive at this site (ABE-136); however the source was never successfully localized. Given the corroborating evidence between Phase-1 and Phase-2 dives at this site, it is unlikely that these detections represent false alarms. Rather, it suggests incorrect interpretation of the likely source of these detections, or simply an unlucky Phase-3. In some cases, notably ABE-151, BP detections indicate unconfirmed additional vent sites that were not explored on subsequent Phase-3 dives.

In retrospect, a strategy based exclusively on BP detections nearly guarantees successful localization of those vent sites whose BPs were intercepted; however, this strategy may also be too conservative with respect to declaring worthwhile targets for pursuit. This is especially true for Phase-1 dives, where direct BP detection can be quite rare on account of wide trackline spacing. Comparing Phase-1 dives ABE-131 (ELSC, Site 5) and ABE-150 (SMAR) illustrates the tradeoff. No BPs were detected on either dive. Nevertheless, eH data from ABE-131 led to two subsequent dives at this site, neither of which ultimately discovered any further evidence of BP activity,

and the site was eventually abandoned.⁴⁸ Additional Phase-1 dives at this site would likely have been a better expenditure of time. In contrast, and despite very wide trackline spacing (1 km), ABE-150 led to the successful discovery of two distinct vent sites, and Phase-2 dives indicated several others in the vicinity. Success was achieved at SMAR largely because eH measurements constrained the younger parts of the NBP. Even so, significant eH anomalies were observed over a large portion of ABE-150. The first of two Phase-2 dives was designed to cover an area beneath the strongest observed eH anomalies. Though successful, both the vent fields discovered during this dive that were eventually confirmed by Phase-3 dives were discovered near the edges of that survey, and could have easily been missed completely. Survey design would likely have benefitted from incorporating measurements of ADCP-derived crossflow velocity.

Efficiency: Relative to the uncertainty in source location derived in this chapter, most Phase-2 and particularly Phase-3 dives could have been designed to cover smaller areas⁴⁹ *if the goal were only to follow up on BP detections*. For instance, at ELSC Site-3 (Fig. F-17), all vent sites eventually confirmed were first detected on the Phase-1 dive at this site. The Phase-2 dive could have been considerably smaller in extent without missing any of these vent sites; however, it was not known at the time whether other BP within the survey area might have been missed. This chapter developed constraints on source location following detection of a BP, but those constraints, while useful, do not provide any guidance on how likely BP detection events are in the first place. One of the reasons for pursuing a nested survey approach to search is to guarantee coverage of a predetermined area. Any measure of performance must include an assessment of how well that goal was achieved. As the next chapter reveals, the information contained in non-detections provides the key to assessing coverage.

Autonomy: One necessary component for the automation of nested surveys for hydrothermal vent discovery has been satisfied: autonomous BP detection. A strategy for survey design based upon those detections remains unspecified. Autonomous

⁴⁸Venting was discovered by a subsequent expedition nearby to the area where the ABE dives took place [50].

⁴⁹In some cases extenuating circumstances also drove survey design. In particular, bathymetry acquired during Phase-2 sometimes indicated geological features worth investigating.

nested survey design would require the specification of survey extent, trackline spacing, and survey height. The development in this chapter does provide guidance on survey extent: an obvious heuristic strategy would be to require that surveys cover areas within some number of standard deviations of predicted source locations. However, it remains unclear (1) how densely the tracklines within should be spaced, (2) how information from multiple nearby detections might be aggregated, and (3) whether the surveyed areas could be declared searched with some measure of certainty. Occupancy grid mapping, applied to the data from this chapter in the next, provides one mechanism to answer these questions.

Chapter 5

Buoyant Hydrothermal Plume Source Localization via Nested Survey

I have pursued a map-based approach toward nested survey automation. The final components in this approach are the occupancy grid (OG) maps of probable seafloor vent location presented in this chapter. These maps are sufficient to provide guidance on both the location and extent of a subsequent survey because they encapsulate all sensor data from the previous stage(s) into a simple spatial representation of seafloor area likely to contain active hydrothermal vents.

5.1 Introduction

In the previous chapter a simple model for the location of a buoyant particle emitted by a vent within a source field of known location was derived. After accounting for measurement uncertainties, this led to a probability density function for predicted particle location. Given particle location instead of source field location, it was remarked that this same probability density function applied to the now uncertain source location.¹ In effect, a forward perspective was utilized to arrive at an inverse model for mapping the locations of BP detections to their likely sources on the seafloor.

¹In fact, this duality applies only for a flat bottom.

However, the inverse model so attained suffers several shortcomings. First, it applies to detections only; the vast majority of measurements are non-detections that consequently cannot be utilized. Second, it offers no guidance on how repeated detections might be combined to yield refined estimates of source location, particularly when multiple sources could be present in the domain. Both these shortcomings can be addressed by retaining the forward perspective and using the OG methods developed in Part I to iteratively construct a map of probable source locations consistent with all measurements.

5.2 Implementation

Figure 5-1 schematically depicts the role of OG mapping within the larger context of nested survey applied to hydrothermal vent localization. The diagram shows how the components of this thesis work together:

1. Start with a survey design informed by the model of the phenomenon being surveyed (§ 4.5).
2. Sensor data collected along the vehicle trajectory is distilled into binary detections and non-detections (§ 4.3). Sensor data may also be used to specify elements of the forward model if it includes dynamic elements (e.g. crossflow velocity).
3. The detection record and forward model are incorporated into probabilities of source occupancy on a discrete grid via an OG mapping algorithm (Ch. 2 & 3). If only a single stage survey was specified, the resultant map is the final output.
4. Otherwise, the next nested survey is designed based upon the state of the OG map at the conclusion of the current survey stage (this chapter).

I concentrate in this section on specifying the internal workings of the elements within the core mapping portion of Fig. 5-1. In particular, I reformulate the model for source

location uncertainty developed in § 4.5 as a binary forward sensor model for use with the OG methods developed in Part I of this thesis, describe the subsampling required to satisfy the “static world” assumption (§ 2.1.1), and discuss setup of the OG map itself (bathymetry, grid cell size, and prior \bar{P}_c^0).

5.2.1 Forward Model for BP Detection

The OG methods of Part I were formulated for use with a specific kind of forward model requiring specification of the quantities P_s^t for all grid cells $s \in \{1, \dots, C\}$. Recall that the quantity P_s^t denotes “the probability that sufficient signal from occupied cell s arrived at the sensor at time t to trigger a detection” (§ 2.3). This section derives an expression for the P_s^t from the particle location probability density function $p(\mathbf{x}_i \mid \bar{\mathbf{x}}_s; h_{vs})$ derived in the last chapter. As will be shown, two additional quantities are required: (1) the probability of detecting a particle given its location relative to the sensor, and (2) the number of such particles composing the plumes emitted from a source field.

Recall from § 4.5 that, given a vent field’s centroid location $\bar{\mathbf{x}}_s$, particles released from the field will be distributed in the water column according to

$$p(h\mathbf{x}_i \mid h\bar{\mathbf{x}}_s; \bar{z}_s) = \mathcal{N} \left\{ h\bar{\mathbf{x}}_s + h\mathbf{U}_0 \frac{h_{\text{CL}}}{W_0}, \left(\left(\frac{1}{2}ah_{\text{CL}} \right)^2 + \sigma_s^2 \right) \mathbf{I}_{2 \times 2} \right\} \quad (5.1a)$$

$$z_i(h_{\text{CL}}) = \bar{z}_s - h_{\text{CL}}. \quad (5.1b)$$

To relate this probability density to the probability of detecting particle i , particles must have some characteristic size. Let each the particle position $h\mathbf{x}_i$ be redefined to instead denote the center of mass of a finite size parcel of effluent having some characteristic radius $b(h_{\text{CL}})$ in the horizontal plane. Let the detection of parcel i at time t be denoted d_i^t and suppose that detection will occur if the detector is immersed within the confines of the parcel. Taking $h\mathbf{x}_i^t$ to denote the horizontal location of the centroid of parcel i at time t , the probability of parcel i triggering a detection at time t is

$$P[d_i^t \mid h\bar{\mathbf{x}}_s; \bar{z}_s] = \begin{cases} P[|h\mathbf{x}_i^t - h\mathbf{x}_v^t| < b(h_{vs}^t) \mid h\bar{\mathbf{x}}_s], & h_{vs}^t = \bar{z}_s - z_i^t \\ 0, & \text{otherwise} \end{cases} \quad (5.2)$$

Larger parcels (up to plume width) are more likely to be detected than smaller ones

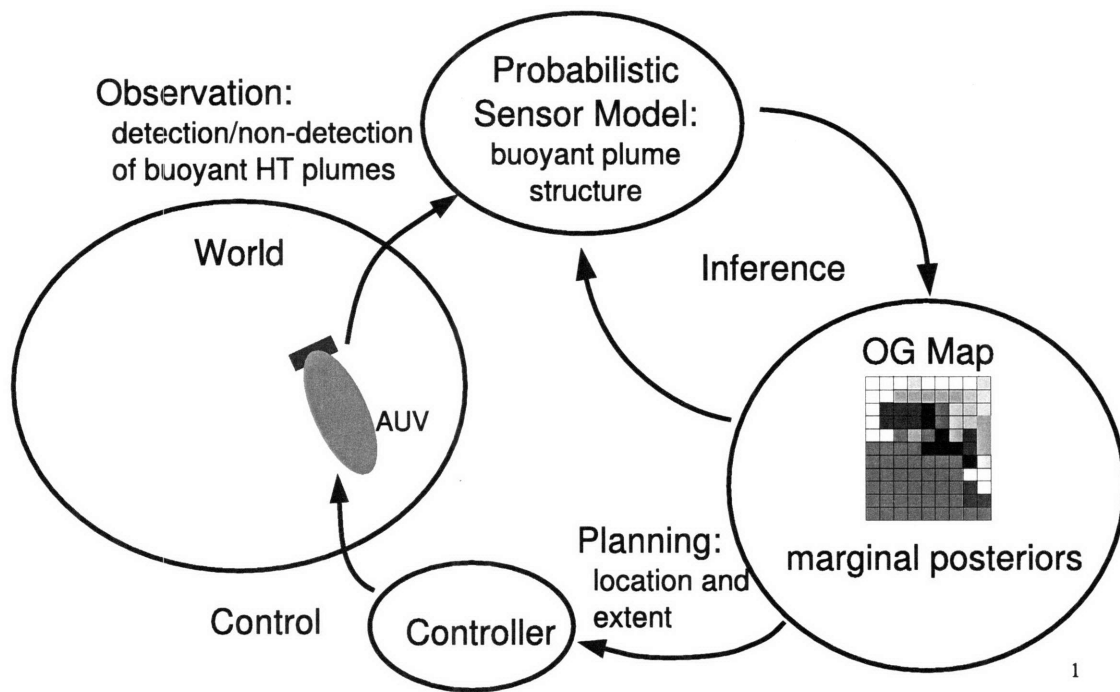


Figure 5-1: Occupancy grid map-driven nested survey applied to hydrothermal vent localization. The AUV inhabits a search domain and flies the first, predefined stage of it's nested survey while constructing an OG map of likely seafloor vent locations. The mapping process consists of acquiring sensor data, interpreting those data via a probabilistic forward model of source location, and applying an OG mapping algorithm to iteratively learn the state of an OG map. Upon completion of one stage of the nested survey, regions of the OG map with posteriors raised above the prior define the location and extent of the follow-on stage. The OG representation facilitates closure of the loop depicted by virtue of its simple indication of coverage (cells with posteriors beneath the prior) and of regions likely to yield additional data useful to refine source location (cells with posteriors raised above the prior).

simply because the vehicle is more likely to intercept a larger target. The probability of immersion in parcel i when $h_{vs}^t = \bar{z}_s - z_i^t$ is approximately

$$\begin{aligned} \text{P} \left[|h_{\mathbf{x}_i^t} - h_{\mathbf{x}_v^t}| < b(h_{vs}^t) \mid h_{\bar{\mathbf{x}}_s} \right] \approx \\ (b(h_{vs}^t))^2 \cdot \frac{1}{2\pi \left(\left(\frac{1}{2} a h_{vs}^t \right)^2 + \sigma_s^2 \right)} \exp \left(-\frac{|h_{\bar{\mathbf{x}}_s} + h_{\mathbf{U}_0^t} \frac{h_{vs}^t}{W_0} - h_{\mathbf{x}_v^t}|^2}{2 \left(\left(\frac{1}{2} a h_{vs}^t \right)^2 + \sigma_s^2 \right)} \right) . \end{aligned} \quad (5.3)$$

This approximation holds as long as patch size is small relative to the uncertainty in parcel location: $b(h_{\text{CL}}) \ll \left(\left(\frac{1}{2} a h_{\text{CL}} \right)^2 + \sigma_s^2 \right)^{1/2}$.

Thus far, $h_{\bar{\mathbf{x}}_s}$ has been regarded as continuous-valued. Assuming a sufficiently fine grid $\Delta \ll \sigma_s$, where Δ denotes the length of the sides of each grid cell, the probability of detecting parcel i given source location is approximately equal to the probability of detecting that parcel given that its source lies within cell s :

$$\text{P} \left[d_i^t \mid h_{\bar{\mathbf{x}}_s}; \bar{z}_s \right] \approx \text{P} \left[d_i^t \mid \mathcal{S} = s; \bar{z}_s \right] \quad (5.4)$$

where the single occupied cell s is chosen such that $h_{\bar{\mathbf{x}}_s}$ lies within its confines. The approximation in (5.4) applies to any map with cell s occupied and not just to maps \mathbf{m} with the set of occupied cells confined to a single cell $\mathcal{S} = s$. However, the rest of the map is completely irrelevant to the probability on the left hand side because parcel i is inherently associated with a particular source field. The right hand side of (5.4) reflects simply a convenient choice of map; convenient because $\text{P} \left[d_i^t \mid \mathcal{S} = s; \bar{z}_s \right]$ happens to be related to the total probability of detection a-specific to any particular parcel.

To get at the total probability of detection, all parcels released by a source field must be considered together. It does not matter which specific parcel is detected, only that at least one parcel is detected. The probability of detection within a map having only one occupied cell is given by

$$\text{P} \left[d^t \mid \mathcal{S} = s \right] = 1 - (1 - P_F^t) \prod_i (1 - \text{P} \left[d_i^t \mid \mathcal{S} = s \right]) , \quad (5.5)$$

where the product is over all parcels released by the source and I have dropped the dependence on bathymetry \bar{z}_s for convenience. For a continuously releasing source,

the product in (5.5) has an infinite number of terms; however, with the vertical position of each parcel z_i deterministic, most of these terms are unity so that:

$$P [d^t | \mathcal{S} = s] = 1 - (1 - P_F^t) \prod_{i: z_i^t = \bar{z}_s - h_{vs}^t} (1 - P [d_i^t | \mathcal{S} = s]) . \quad (5.6)$$

Furthermore, $P [d_i^t | \mathcal{S} = s]$ is identical for all parcels i satisfying $z_i = \bar{z}_s - h_{vs}^t$. This yields the particularly simple expression

$$P [d^t | \mathcal{S} = s] = 1 - (1 - P_F^t) (1 - P [d_i^t | \mathcal{S} = s])^Q . \quad (5.7)$$

where Q denotes the number of parcels present from the source in cell s at vehicle depth. The number Q is analogous to a release rate and as such should therefore be constant with height above the seafloor.

The final step to attaining an expression for the P_s^t requires relating the P_s^t to the single-occupied-cell map via (2.12), repeated here for completeness:

$$P [d^t | \mathcal{S} = s] = 1 - (1 - P_F^t)(1 - P_s^t) . \quad (5.8)$$

Substituting the results from (5.7) and (5.3) into the above yields

$$\begin{aligned} P_s^t &= 1 - (1 - P [d_i^t | \mathcal{S} = s])^Q \\ &= 1 - \left(1 - (b(h_{vs}^t))^2 \cdot \frac{1}{2\pi \left(\left(\frac{1}{2} a h_{vs}^t \right)^2 + \sigma_s^2 \right)} \exp \left(- \frac{|h \bar{\mathbf{x}}_s + h \mathbf{U}_0^t \frac{h_{vs}^t}{W_0} - h \mathbf{x}_v^t|^2}{2 \left(\left(\frac{1}{2} a h_{vs}^t \right)^2 + \sigma_s^2 \right)} \right) \right)^Q \end{aligned} \quad (5.9)$$

It remains to specify the parameters $b(h_{CL})$ and Q . Those estimates are the subject of the following two short sections. Once estimates for parameters have been attained, the forward model required by the OG methods of Part I will be completely specified.

Parcel population An appropriate choice for parcel population Q depends primarily on the accuracy of the plume model employed. The rather large characteristic source size $\sigma_s \approx 25$ m found in Ch. 4 indicates our model is insufficiently accurate to predict the location of individual vents on the seafloor, instead being suited to

predicting the locations of entire vent fields. Parcel population Q should therefore reflect the number of discrete plumes emanating from a field. Unfortunately such knowledge is obviously not available a priori. A “reasonable” value, say $1 \leq Q \leq 5$ unless especially large vent fields are expected, is the only viable recourse.

Parcel size Figure 5-2 shows encountered patch size as a function of vehicle height above bottom at the time of interception. Patch size was determined by first temporally clustering full sample rate detections and then computing the maximum pairwise distance between detections within each group. Almost all patches were continuous at the sample rate of ~ 1 Hz, though some were bimodal. Most patches at low altitude substantially exceed the expected width of a point-source plume assuming a linear spreading rate of $a = 0.1$ m/m. This could reflect interception of multiple plumes within a single patch; however, even some of the largest patches were unimodal. Of course, on the short timescale of a single transect through a buoyant plume, the observed profile cannot be expected to reflect time-averaged behavior. Based on Fig. 5-2, a linear growth rate would significantly under-predict of the probability of plume interception at 50 m. Instead, I employ an affine model for average plume size as a function of centerline rise height h_{CL} :

$$b(h_{CL}) = b_0 + \frac{1}{2}ah_{CL} . \quad (5.10)$$

The constant b_0 represents the effect of finite source size. The height dependence was chosen to be consistent with the assumptions of the particle model for plume growth (4.23). Unfortunately, Figure 5-2 provides little guidance on the correctness of this choice, particularly because of the dearth of data at large heights. The $\frac{1}{2}ah_{CL}$ line for $a = 0.1$ is plotted on Fig. 5-2 and suggests at least plausibility, especially considering that patch width at interception will tend to be somewhat smaller than maximum patch width.

A maximum likelihood estimate of b_0 for each dive was attained from groundtruth

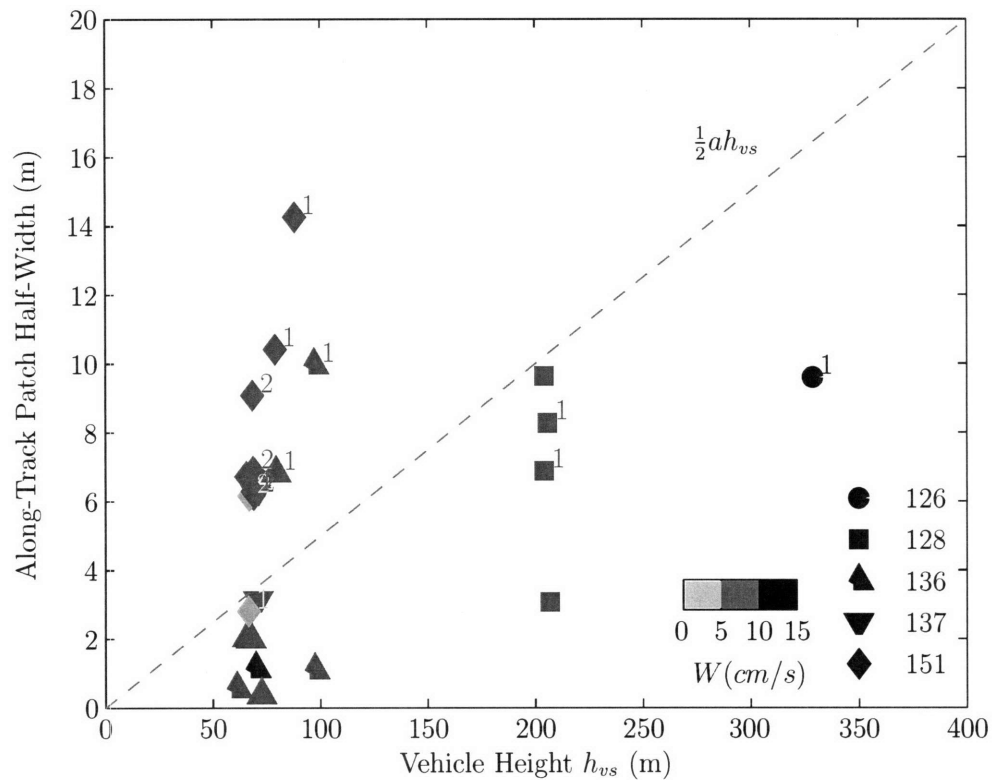


Figure 5-2: Observed plume width vs. vehicle height h_{vs} . Widths plotted are one-half the portion of trackline associated with each detection. Superscripted numbers indicate detections that clearly showed either 1 or 2 modes in vertical velocity. Observed widths were broadly distributed but indicate much larger average widths at $h_{vs} \sim 50$ m than would be expected from point sources with linear radial growth rates of 0.1 m/m. This observation motivated the affine model proposed in the text.

Table 5.1: Maximum likelihood estimates of characteristic source size b_0 for each dive that included interception of BP water.

Dive Number	Dive Type	Location	\hat{b}_0
ABE-126	Phase-1	ELSC-1	12
ABE-128	Phase-1	ELSC-2	6
ABE-136	Phase-2	ELSC-2	4
ABE-137	Phase-2	ELSC-1	4
ABE-151	Phase-2	SMAR	13

vent field location and measurement history Δ^t according to:

$$\begin{aligned}
\hat{b}_0 &= \operatorname{argmax}_{b_0} P[\Delta^t \mid \mathbf{m}] \\
&= \prod_{\tau=1}^t P[\delta^\tau \mid \mathbf{m}] \\
&= \prod_{\tau \in \tau^t} P[d^\tau \mid \mathbf{m}] \cdot \prod_{\tau \in \bar{\tau}^t} (1 - P[d^\tau \mid \mathbf{m}]) ,
\end{aligned} \tag{5.11}$$

where the second line follows from the static world assumption (§ 2.1.1). The probabilities appearing in the last line of the above are given by the forward model (2.11) with the P_s^t from (5.9). Numerical estimates were attained using all groundtruth vent locations (with $Q = 1$ for each) and are listed in Table 5.1.

The values for b_0 in Table 5.1 show considerable variation between dives. This is not unexpected, as the estimate will depend strongly on vent field configuration and each individual detection record. Furthermore, these estimates do not take into account that crossflows cannot be known exactly, and so will tend to produce overestimates of b_0 . Ultimately a choice must be made that is reasonable across all dives. The occupancy grid maps produced later in this section were produced using $b_0 = 5$ m.² Increasing b_0 increases the likelihood of detections at smaller heights above seafloor proportionally more so than at larger heights.

In light of a characteristic parcel size specified by (5.10), the accuracy of the single-point integration scheme in (5.3) might be questioned. The worst case corresponds

²It is likely that the value for ABE-126 above is high on account of the large height of this dive (300 m) meaning that crossflow uncertainty will have corrupted the estimate.

to the largest height above seafloor of interest. For the parameter values used in this thesis, characteristic parcel size at 300 m above the seafloor will be 20 m while the standard deviation associated with parcel location will be about 40 m. Though only separated by a factor of two, single-point integration continues to provide a reasonable answer in practice.

5.2.2 Input Data

The OG mapping algorithms developed in Part I of this thesis require a binary input: detection or non-detection. Distillation of the raw sensor data collected on ABE into binary measurements was discussed in § 4.3; however, the development applied to data at full sample rate ~ 1 Hz. Solitary detections at this sample rate have yet to be observed (cf. patch width data presented in Fig. 5-2) indicating samples at ~ 1 Hz are correlated. Blind use of the detection record at full bandwidth would therefore violate the static world assumption (§ 2.1.1). Recall that this assumption requires that measurements be independent conditioned on knowledge of the entire map \mathbf{m} .³ At full sample rate, and independent of \mathbf{m} , a detection makes it likely that the next measurement will also be a detection simply because hydrothermal BPs have significant width at survey height compared to the spatial bandwidth implied by this sample rate.

To avoid treating correlated data as independent, the full bandwidth detection record consisting of binary samples δ_o^t is downsampled. The anti-aliasing applied is non-standard in the sense that the output δ^t must still be binary. This is done by declaring downsampled detections when at least a fraction $\eta \in [0, 1]$ of the original samples within the new sampling period were detections:

$$\delta^t = \left[\left| \{ t : \delta_o^t, t \in [t - T/2, t + T/2) \} \right| > \frac{T}{T_o} \eta \right] \quad (5.12)$$

where T_o denotes the original sampling period, T denotes the downsampled period,

³The static world assumption does not require that the entire world actually be static. Crossflow velocity is readily incorporated into specification of the probabilities of detection P_c^t ; concentration of tracers with known decay rates could similarly be incorporated into this specification by raising the P_c^t for cells at the proper upwind distance from the sensor. On the other hand, plume sources that move or periodically turn off cannot be accomodated without violating the static world assumption because in this case the true OG map would become time-dependent.

and the enclosing square brackets imply a boolean result. An $\eta = 0$ makes every interval containing at least one detection at the original sampling frequency into a detection at the new sampling frequency. An $\eta = 1$ requires that all measurements at the original sample rate within the interval be detections before declaring a detection at the new sampling frequency. Large η close to 1 decrease the probability of false alarm at the expense of decreased probability of detection. The results in § 5.3 were all produced with $\eta = 0.5$.

Choosing a good downsampling period T is crucial. Large T will decorrelate the samples more effectively, but reduce the precision of the measurement's location. A good choice for T will reflect the spatial scales of the phenomenon of interest. For BP detection, T times vehicle speed $|\dot{\mathbf{x}}_v|$ should be on the order of expected plume width at survey height:

$$T \approx \frac{1}{|\dot{\mathbf{x}}_v|} (b_0 + \frac{1}{2}ab) . \quad (5.13)$$

Thus for a Phase-2 survey at 50 m height and ABE's typical survey speed of 0.6 m/s, $T \sim 10$ s. At 300 m height, $T \sim 30$ s.

5.2.3 Mapping Domain

For hydrothermal vent mapping, a grid is defined on the seafloor. Bathymetry acquired from a surface ship will have a resolution on the order of 100 m, which is adequate. Grid cell size and a prior probability of occupancy need to be specified. The forward model formulated above applies to the locations of entire vent fields. Grid cell size should be small relative to the characteristic uncertainty associated with source location. The lower bound is set by σ_s , about 25 m in the data sets studied. A grid cell size of $\Delta = 5$ m on a side is sufficiently dense in practice. At larger heights, a lower resolution is sufficient and also desirable because these dives typically span a larger area and computational load grows quadratically with survey area (linearly with the number of grid cells).

An appropriate cell prior reflects the expected number of vent fields in the survey area scaled by grid resolution. Let ρ_f denote the expected the field density, that is, the expected number of fields divided by the survey area: $\rho_f \triangleq \text{E}[S] / (C\Delta^2)$.⁴ The

⁴Recall from Ch. 2 that \mathcal{S} denotes the set of occupied grid cells in a map and C denotes the number of grid cells in the map.

expectation evaluates to $E[S] = C\bar{P}_c^0$ so that the prior probability of occupancy for each cell is given by:

$$\bar{P}_c^0 = \Delta^2 \rho_f . \quad (5.14)$$

All OG maps presented in this work employed an assumed vent field density of $\rho_f = 1 / \text{km}^2$, yielding priors between about 10^{-5} and 10^{-3} depending on grid cell size.

For nested surveys, the finished OG map from one stage can be used to initialize a higher resolution OG map for the subsequent stage. To do so requires regarding the final posteriors of the original map as independent. With this assumption, the new priors can be attained from the old posteriors by dividing the old posteriors equally among each of the new grid cells within the corresponding cell of the old grid such that expected occupancy of the region contained within the original cell is unaltered.⁵ This transformation also preserves the expected number of source fields from the finished map. Finally note that the assumption of independence required is consistent with the assumptions required by the IP algorithms developed in § 3.2.

5.2.4 Choice of OG Algorithm

Figure 3-10 illustrates the factors affecting choice of OG algorithm. The low prior appropriate for hydrothermal vent field localization rules out the standard algorithm. This leaves the exact algorithm and the various approximations developed in Ch. 3. Since only very few BP detections (about 10 or less, and distributed between different vent fields) are expected during each survey stage, the exact formulation could have been applied; however, the recursive form of the IP algorithm is the most practical from the perspective of real-time application and the results presented next all employ the IP approximation. Only minor degradation of the IP maps relative to the exact result was observed. Appendix G contains a thorough comparison to the exact result along with an example of the highly degraded map produced by the standard algorithm applied to the same data.

⁵Kraetzschmar et al. [59] use this same “probabilistic mean” to propagate probabilities of occupancy through a multi-resolution probabilistic quad tree.

5.3 Results: Vent Field OG Maps

Figures 5-3 & 5-4 show OG maps generated from data collected on two separate dives, ABE-128 (Phase-1) and ABE-136 (the subsequent Phase-2 dive). Both detections and non-detections were used as input data to the recursive form of the IP algorithm (§ 3.2.1). Regions with relatively high posterior odds of occupancy (hot colors) agree well with clusters of groundtruth vent locations. Regions with relatively low posterior odds (cool colors) indicate coverage over regions where no BPs were encountered. Multiple detections plus nearby non-detections create the irregular shape of some of the regions with relatively high posterior odds.⁶ However, the influence of non-detections is relatively weak because detections are unlikely according to the forward model of § 5.2.1 and consequently non-detections carry less information than detections.

Figure 5-4 illustrates the result of using the posterior from a previous dive as a prior. ABE-128 and ABE-136 were conducted at ~ 250 m and 50 m above bottom respectively. Because of this difference in height, the grid cell sizes chosen were different: 20 m and 5 m on a side respectively. Each cell in the map shown in Fig. 5-3 was split into sixteen descendant cells, each of which were initialized with a probability of occupancy $1/16$ that of the posterior in the original parent cell, as per the procedure in § 5.2.3. In this case, incorporation of the Phase-1 result improved the resulting map by retaining high posteriors over portions of the fields that were not detected during ABE-136.

Both maps were produced using the recursive form of the IP OG mapping algorithm (§ 3.2.1). The number of detections was sufficiently small to allow use of the exact algorithm (§ 2.4), however, the resulting maps are qualitatively similar, and the recursive IP algorithm completely avoids the numerical and computational vulnerabilities of the exact algorithm. Furthermore, propagating the posterior from previous dives into a higher resolution grid via the procedure in § 5.2.3 is consistent with the assumption of independent posteriors required by the IP algorithm. The standard algorithm (§ 3.1.1) applied to this data produces posteriors near unity in hundreds of cells near some of the vent fields, in gross violation of the assumed low prior but

⁶Using the recursive form of the IP OG algorithm to create these maps instead of the exact formulation also has some effect because, unlike the exact algorithm, the IP algorithm is incapable of reducing any posterior odds following a detection.

consistent with the observed failure mode of the standard algorithm in simulation (§ 3.3). Appendix G contains a cross-comparison between maps produced using all three algorithms.

5.3.1 Conversion from Field Map to Vent Map

The forward model developed in § 5.2.1 applies to the probability of detection given knowledge of vent field location. The posterior odds in OG maps like Fig. 5-3 & 5-4 therefore indicate the odds that a cell is occupied by the center of a vent field. As long as vehicle height exceeds characteristic vent field size σ_s , this perspective is appropriate. Closer to the seafloor (e.g. Phase-3 dives at 5 m height) it is desirable to map the individual vents within a vent field. To choose the extent of a low-height survey based on an OG map, the cell posteriors should represent the probability that they are occupied by individual vents.

The transformation is straightforward. I assumed in § 5.2.1 that a typical vent field will contain Q individual vents whose locations will be normally distributed about the field's center ${}^h\bar{\mathbf{x}}_s$. Under these assumptions and given ${}^h\bar{\mathbf{x}}_s$, the probability of a cell at ${}^h\mathbf{x}_c$ containing a vent is

$$P[v m_c | \bar{\mathbf{x}}_s] \approx 1 - \left(1 - \frac{\Delta^2}{2\pi\sigma_s^2} \exp\left(-\frac{|{}^h\bar{\mathbf{x}}_s - {}^h\mathbf{x}_c|^2}{2\sigma_s^2}\right) \right)^Q, \quad (5.15)$$

where ${}^v m_c$ denotes occupancy of cell c by a vent. In words, the probability that a cell c is occupied by a vent is one minus the probability that it is not occupied by any of the Q vents in the field. Multiplication by Δ^2 approximates integration over the area of cell c . Now let ${}^f m_s$ denote that cell s is occupied by the center of a vent field. Supposing $\bar{\mathbf{x}}_s$ lies within a cell s , $P[v m_c | {}^f m_s] \approx P[v m_c | \bar{\mathbf{x}}_s]$. To transform the the posterior probability of field occupancy $P[{}^f m_s | \Delta^t]$ to the posterior probability of vent occupancy $P[v m_c | \Delta^t]$, sum over all possible field locations:

$$P[v m_c | \Delta^t] = \sum_{s=1}^C P[v m_c | {}^f m_s] P[{}^f m_s | \Delta^t]. \quad (5.16)$$

Figure 5-5 shows the results of applying this transformation to the field OG map of Fig. 5-4.

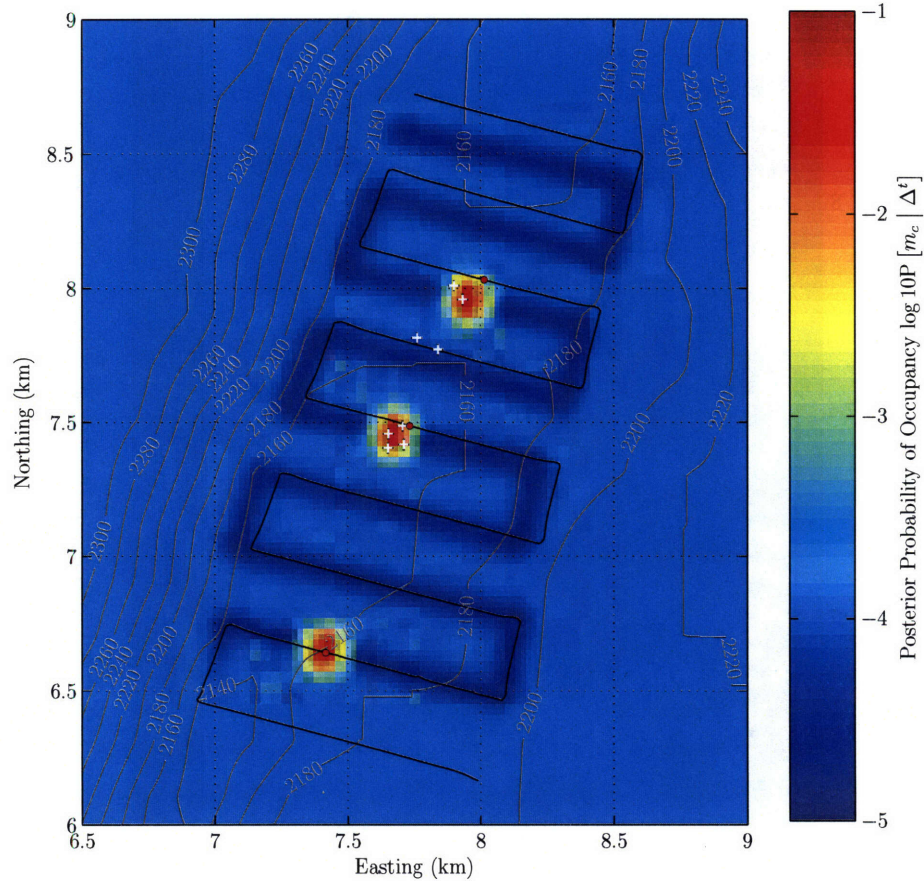


Figure 5-3: OG Map produced from Phase-1 dive ABE-128. Groundtruth vents are shown as white crosses. The vehicle registered detections near two of the three vent fields ultimately localized in addition to a detection near the southern end of the survey, the source location of which remains unknown. High posterior odds in the occupancy grid map (red) show good agreement with the locations of the two detected vent fields. Lowered posterior odds (dark blue) indicate trackline spacing was insufficient by roughly a factor of two. In fact, the dive shown included a first pass over this site at a different depth and offset by half the trackline spacing shown, which would have doubled the trackline density except that vehicle depth was shallower than the NBP making BP contact unlikely. A subsequent survey designed purely based upon this map would have discovered two of the three vent fields shown and possibly a third site to the south, though it would have missed one field. However, that field was missed because it was not detected, and not because the OG algorithm failed to produce a good map.

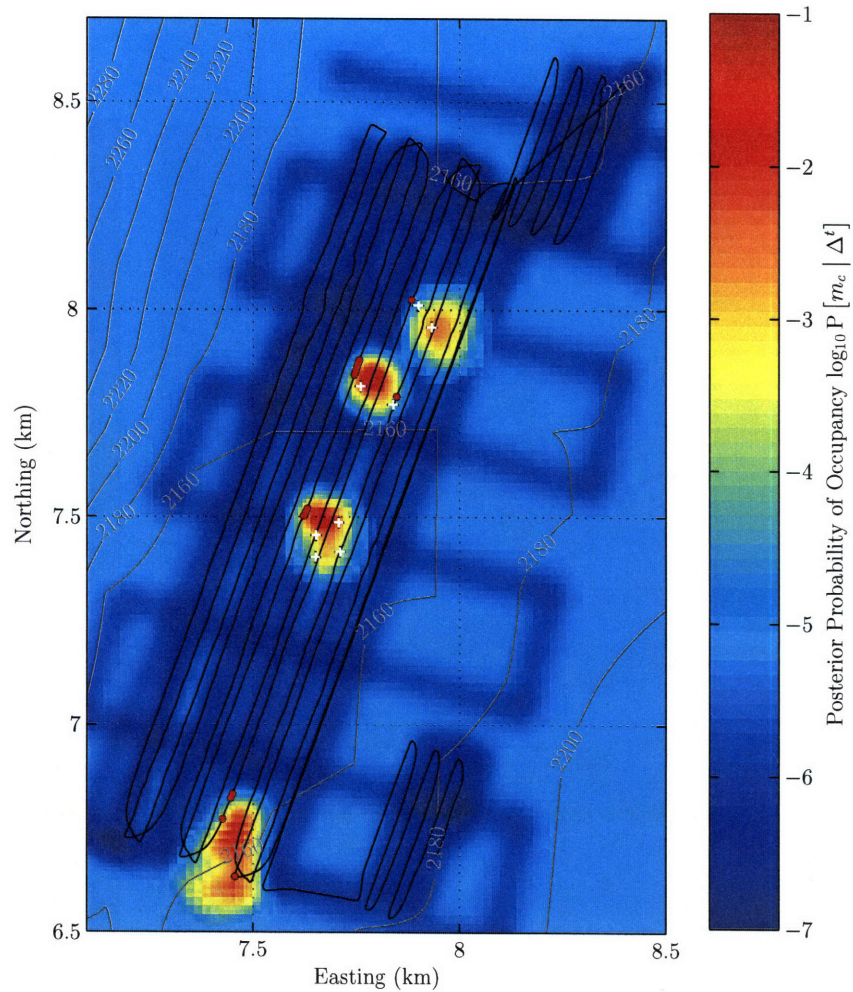


Figure 5-4: OG Map produced from Phase-2 dive ABE-136 using the map of Fig. 5-3 to define the prior. The locations of all vent fields ultimately confirmed agree well with the regions of the map populated by cells with high posterior odds (hot colors). Detections at the southern edge of the survey confirm the presence of an additional field, first detected during the Phase-1 dive at this site; however, this southern site was never groundtruthed. Regions with low posterior odds (cool colors) indicate fairly uniform coverage over most of the survey extent, though with some patchiness particularly in the southern half of the survey area. Interestingly, this site was surveyed in two passes, first progressing to the west, then back toward the east in between the previous tracklines. The crossflow direction was observed to approximately reverse over the course of the survey, resulting in tracklines that effectively observed the same portion of the seafloor as the previous pass. When these portions of the seafloor included vent sites, the data was valuable to “triangulate” the location of the vent field. On the other hand, when these portions of the seafloor were empty, little additional information was attained. In retrospect, improved coverage would have been attained by a single pass at full trackline density.

The transformed map shows broader peaks over wider areas than the field location map in Fig. 5-4, a result that is consistent with vent fields of finite size. Also note that the residual high-frequency features in the map of Fig. 5-4 left over from incorporation of the lower resolution prior have been smoothed. Maps like these could be used to guide subsequent near bottom Phase-3 mapping surveys of individual vent fields.

5.3.2 Survey Design from OG Map

The algorithm used to generate the map of Figs. 5-5 incorporates tracer data and measured crossflows into consistent maps of the seafloor that identify both regions worth revisiting and regions unlikely to contain vents. Following completion of one stage of a nested survey, design of the next stage should favor tracklines that provide coverage over the portions of the map raised above the prior. Likewise, time spent over portions of the survey area with lowered posteriors will likely be wasted fruitlessly, and so should be avoided. For example, Fig. 5-5 includes the superimposed outline of a Phase-3 dive, that while successful in acquiring groundtruth locations for 5 of 8 vents ultimately discovered, also included substantial coverage of seafloor unlikely to contain vents according to the OG map. Dive time could have been allotted more efficiently, and probably would have yielded additional groundtruth vent locations.

Figure 5-6 presents a quantitative evaluation of the efficiency of Phase-2 dives relative to OG maps from preceding Phase-1 surveys and of Phase-3 surveys relative to OG maps from preceding Phase-2 surveys. The figure indicates most follow-on surveys covered too much ground unlikely to contain vents. In many cases, potentially high value areas were consequently ignored even though they could have been surveyed without extending the duration of the dive.⁷ On the other hand, wider coverage sometimes resulted in detections of vent fields that would have been missed had subsequent surveys been designed purely based upon the OG map of the previous stage. For instance, the second field from the north in Fig. 5-4 is not evident in the OG map from the previous dive Fig. 5-3. However, in these cases, either insufficient coverage from the previous survey or else a missed detection was to blame. The BP detection method implemented in § 4.3 was deliberately designed with a very low

⁷Other factors did sometimes affect survey design, for instance, bathymetry, CTD data, and imposed time-constraints.

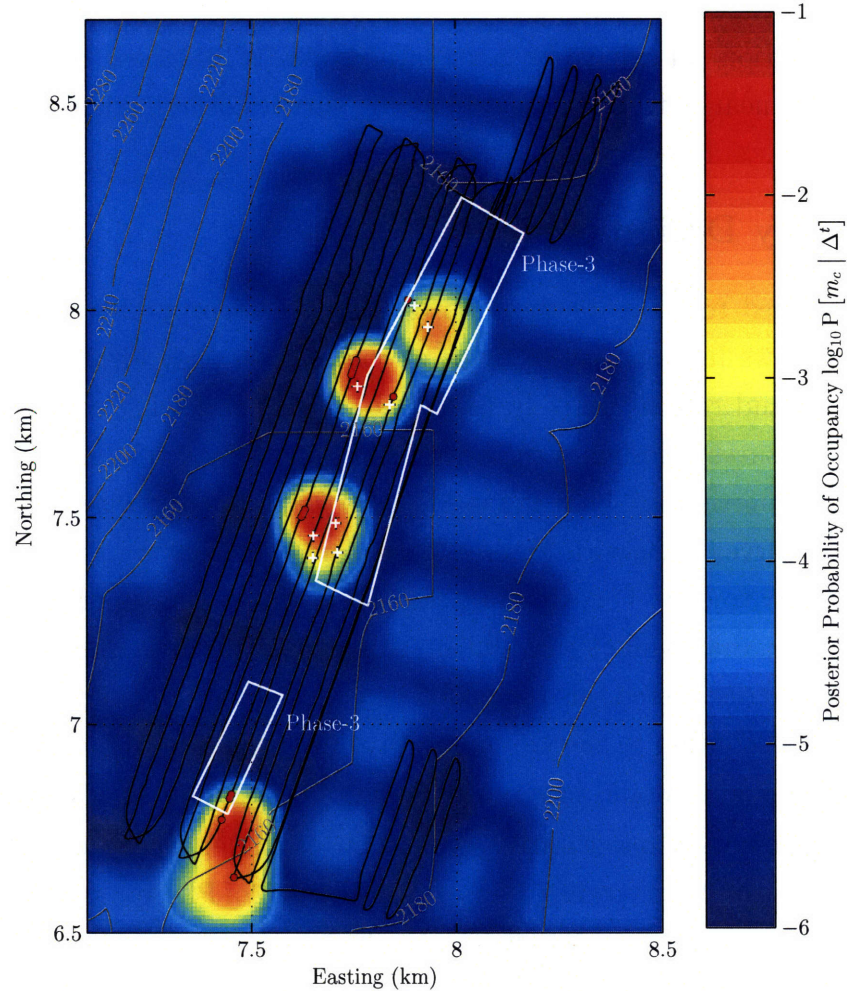


Figure 5-5: The OG map of Fig. 5-4 transformed from probability of vent field occupancy into individual vent occupancy. Relative to the field map, this map exhibits wider, smoother peaks consistent with vent fields of significant extent (10s of meters). The outline of the subsequent Phase-3 near-bottom survey at this site is shown in white. That survey was designed while at sea on the original expedition, before development of the OG mapping framework. Better coverage could have been attained by decreasing the north-south extent of the northern survey in favor of extending it to the west. Indeed, the three western groundtruth vent locations that lie outside the bounds of the Phase-3 survey were discovered by ROV several months after the completion of ABE operations at this site. Finally, the map indicates the southern Phase-3 survey was misplaced, and should have been centered some 200 m to the southeast.

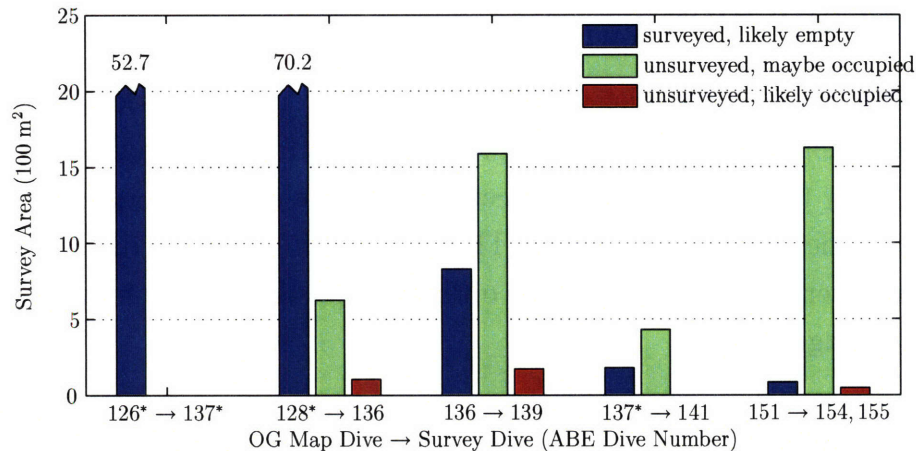


Figure 5-6: Survey coverage relative to high and low value map areas identified in OG maps. The labels indicate the dive used to create the OG and the dive whose coverage is evaluated relative to that map: map dive → subsequent survey. For each pair of dives, three areas were computed: (blue) area covered by the subsequent survey but less than the prior ($< 0.99\bar{P}_c^0$); (green) area not covered but raised above the prior ($> 1.01\bar{P}_c^0$); (red) area not covered with relatively high posteriors greater than 0.01 (i.e. at least a 1:100 chance of vent occupancy). Qualitatively these areas indicate, respectively, unnecessary coverage, possible missed opportunities, and likely missed opportunities. For the same dive duration, better survey performance might have been attained in most cases by covering the areas represented by the red and green bars instead of the areas represented by the blue bars. Dives that discovered vent fields in places indicated by the OG map as unlikely to have been occupied are possible exceptions. These dives are starred. In all cases, indications of BP contact (usually eH) were present in the data used to make the OG map; however, the evidence was insufficiently compelling (usually lacking significant VVA) to result in a detection according to the method of § 4.3.

probability of false alarm because the forward model employed is specific to direct BP contact.

From the perspective of automating nested surveys, the results in Fig. 5-6 are significant because they indicate that relatively short-duration subsequent stages would have been adequate to localize many vents. Short-duration, vehicle directed stages could therefore have been added to the ends of Phase-1 dives without significantly impacting the overall coverage attained. In many cases, this would have enabled the vehicle to return from a single dive with complete preliminary assessment of at least one new vent site including groundtruthed location, nearby bathymetry and photographs of the surrounding biological community.

5.4 Conclusions

This chapter presented a parameterized forward model for BP detection and applied it to the construction of OG maps of likely vent locations from real world hydrographic data collected by the ABE vehicle on recent hydrothermal vent prospecting missions in both the Pacific and Atlantic oceans. These maps can be used to distill sensor data to aid human interpretation or, by virtue of their simplicity, to drive fully autonomous survey stage design.

Chapter 6

Conclusion

This thesis presented a stochastic mapping framework designed to enable a robotic platform to autonomously localize chemical plume sources in environments with multiple sources. Potential applications for robotic chemical plume source localization include pollution and environmental monitoring, chemical plant safety, search and rescue, anti-terrorism, narcotics control, explosive ordinance removal, and hydrothermal vent prospecting. Turbulent flows make the spatial relationship between the detectable manifestation of a chemical plume source, the plume itself, and the location of its source inherently uncertain. Search domains with multiple sources compound this uncertainty because the number of sources as well as their locations are unknown.

The framework is an adaptation of occupancy grid mapping wherein the binary state of map nodes is redefined to denote either the presence (occupancy) or a absence of an active plume source instead of the usual presence or absence of a physical obstacle. A key characteristic of the chemical plume source localization problem is that only a few sources are expected within the search domain. The occupancy grid framework allows for both plume detections and non-detections to inform the estimated state of grid nodes in the map, thereby explicitly representing explored but empty portions of the domain as well as probable source locations. However, sparsity in the expected number of occupied grid nodes strongly violates a critical conditional independence assumption required by the standard Bayesian recursive map update rule. While that assumption makes for a computationally attractive algorithm, in our application it results in occupancy grid maps that are grossly inconsistent with the assumption of a small number of occupied cells. To overcome this limitation, sev-

eral alternative occupancy grid update algorithms were presented, including an exact solution that is computationally tractable for small numbers of detections and an approximate recursive algorithm with improved performance relative to the standard algorithm but equivalent computational cost.

Application to hydrothermal plume data collected by the autonomous underwater vehicle ABE during vent prospecting operations in both the Pacific and Atlantic oceans verified the utility of the approach. The resulting maps were shown to enable nested surveys for homing in on seafloor vent sites to be carried out autonomously. Real-time implementation would eliminate inter-dive processing, recharging of batteries, and time spent deploying and recovering the vehicle that is otherwise necessary with survey design directed by human operators.

The remaining paragraphs reiterate the contributions made by this work to the three areas identified in Ch. 1: (1) occupancy grid (OG) mapping; (2) general chemical plume source localization; (3) hydrothermal vent prospecting. In each category I identify candidate directions for future research that leverage these contributions.

Occupancy Grid Mapping

- The application of OG mapping to novel arena—multi-source chemical plume source mapping.
- An articulation of the problems associated with application of standard Bayesian OG mapping to environments with few expected occupied cells (low prior probability of occupancy) and stemming from a key independence assumption required by the standard algorithm.
- An exact solution for the state of an OG map when measurements consist of binary detections and non-detections generated by a particular form of forward sensor model.
- A family of novel approximate algorithms applicable to low prior environments based on this exact formulation.

Hybrid Feature-Based and OG Mapping This thesis showed how, for a particular form of binary forward sensor model, non-detections can be incorporated into the map recursively and without approximation. When the number of cells expected to be occupied is low, feature-based mapping methods make for a logical choice because the correspondence problem is low-dimensional. However, feature maps cannot make direct use of sensor information indicating emptiness because these observations cannot be correctly attributed to any one feature in the map, and because feature maps include no explicit representation of empty space. A map of feature locations does not indicate whether a region in the map not containing any features is likely to be empty, or hasn't been searched, or resulted in conflicting information.

A hybrid approach might prove feasible wherein sensor measurements are distilled into binary detections and non-detections, the latter being incorporated into an occupancy grid map, and the former into a feature-based map. For the sum to be greater than its parts, the information in these two maps would have to be linked. For instance, feature locations could be locally biased toward regions less likely to be empty according to the underlying OG map. A hybrid map would be particularly useful in combined mapping/exploration missions in sparsely occupied environments.

Implications for OG mapping in Indoor Environments Low priors are a characteristic of the chemical plume source localization problem when framed as a Bayesian OG mapping problem, an aspect that was shown to exacerbate the deleterious effects of regarding each measurement as conditionally independent of all others given only knowledge of the single cell being updated. Ubiquitous walls and furniture imply that indoor environments (the original application for which OG mapping was developed) are not characterized by low prior probabilities of occupancy when discretized into an OG map. Nevertheless, the assumptions required by the standard OG mapping algorithm do manifest themselves as undue ambiguity around critical features like doorways [117].

It remains to be seen whether the methods developed in this thesis offer any advantages in indoor environments over existing OG algorithms. The particular form of sensor model required can be applied to sonar range-finders (Fig. 2-3; however, the inverse models typically employed in classical OG mapping do not fit the constraints imposed by (2.11). Rather, these highly tuned inverse models (e.g. [116]) tend to

produce good results precisely because they infer more about the environment than is strictly available, at least from the simplified geometric perspective of Fig. 2-3. In particular, indoor environments are structured, that is, they consist of regular walls that imply occupied cells usually occur together, a fact that can be exploited in an inverse model, but not in a forward model wherein each occupied cell can independently trigger a detection at the receiver. Nevertheless, the exact solution presented here enforces consistency between the posteriors and all data, so that in principle its application to indoor environments address the same issues explored by Thrun [117].

Chemical Plume Source Localization

- An abstracted forward model for binary chemical plume detection that encapsulates the role of multiple sources without reference to the physics of a particular type of plume.
- A plume source location mapping method suitable for use in multi-source environments, for instance in mine-clearing, explosive ordinance removal, and in hydrothermal vent prospecting.

Map-Driven Multi-Source Biomimetic CPT In this thesis I focused on the application of OG maps to automating nested surveys. For some types of plumes, nested surveys can be wasteful of vehicle time. In particular, when all sources are equally valuable and detections are unequivocal, completing a survey stage before following up on a detection requires that subsequent surveys span the potentially large portion of the map that may contain the source of that detection. For short-timescale passive plumes, biomimetic strategies provide an attractive alternative, but it remains unclear how best to adapt such strategies to multiple-source domains: a male moth seeking a mate needs only to find one pheromone-emitting female.

Enhanced efficiency could be attained in a multi-source domain over nested surveying by combining biomimetic plume tracing with higher-level mapping to direct the surveying vehicle toward unexplored regions of the survey area and to identify when new detections are likely to have emanated from an already-localized source.

An OG map and associated forward sensor forward model provide a natural mechanism to affect this strategy. The forward model can be employed to generate locations where the plume emanating from a grid node is likely to be detected under the assumption that the node is occupied by a source. The present state of the OG map in turn indicates which cells are most worth attempting to observe. Successful previous localizations will have rendered those portions of the map fully resolved (posterior probabilities of occupancy near one or zero.), whereas cells in unexplored regions will have the most uncertain state and the highest value of observation. One challenge to implementing this strategy is that it will be sensitive to accurate posteriors, particularly if entropy is employed as the metric to assess the relative worth of observing a cell.

Hydrothermal Vent Prospecting

- A procedure for automatic classification of hydrographic data into the background water and the two main components of a hydrothermal plume—the buoyant and non-buoyant plume—and its application to field data.
- An algorithm for the generation of occupancy grid maps of the seafloor showing locations likely to contain hydrothermal vents and also regions unlikely to contain vents based on the data acquired, and suitable for the automation of nested surveys in support of hydrothermal vent prospecting.
- A simple model for buoyant hydrothermal plume evolution suitable for use with the occupancy grid mapping methods developed in this thesis.
- An evaluation of the utility of measuring crossflow velocity on a surveying AUV for constraining the source locations of encountered buoyant plumes.

Real-Time Automated Nested Survey Implementation The OG maps in this dissertation were all generated long after the ABE AUV actually collected the data. Throughout I have tried to be mindful of the constraints imposed by real-time

operation and the often chaotic environment of a research vessel at sea. To these ends, the results presented in Ch. 5 all used the computationally minimal recursive form of the IP OG mapping algorithm, and the same model parameters were used across all dives to simulate ignorance about the specific environment to be encountered on a given dive. Nevertheless, only actual real-time deployment in the deep sea will ultimately prove the utility of these algorithms for automating the nested survey process.

Improving Performance beneath the NBP Reduction-oxidation potential (eH) and potential temperature anomaly data collected during 50 m height Phase-2 dives with the ABE AUV (App. F) indicates vent fields tend to be associated with significant anomalies in these tracers over much larger areas than vertical velocity anomaly (VVA), though the strongest anomalies are well correlated with VVA. The simplistic BP model employed here cannot account for these farther-afield anomalies. Numerical modeling work by Lavelle [66] shows that buoyant hydrothermal plumes rising in crossflows possess an asymmetric potential temperature cross-section in the horizontal plane with significant downwind elongation in the wake of the core of the plume. Presumably a model amenable to real time computation could be developed to capture at least the rudiments of this process (there is no need to accurately predict temperature anomaly for instance). The benefit of including these anomalies in the construction of an OG map to drive nested survey is obvious: trackline spacing could increase without compromising coverage, thereby reducing the time required to survey a prescribed area.

Application to eH Anomalies in the NBP Direct detection of BP stems within the NBP is unlikely when vehicle trackline spacing exceeds the characteristic width of BPs at altitude (< 100 m). Substantial gains in both efficiency and coverage are possible if the age of NBP water can be determined and interpreted in terms of likely BP location without requiring direct BP contact. Emerging in situ chemical sensors, for instance the reduction-oxidation (eH) probe discussed in Ch. 4, promise to be able to provide constraints on the age of NBP water. Indeed, successful vent discoveries at SMAR based on the initial widely (1 km) spaced tracklines of the single Phase-1 dive at the site were almost entirely due to the eH probe providing indications of

relatively young NBP water despite lack of direct contact with any BPs.

Thus far the design of ABE surveys based on eH anomalies in the NBP has been heuristic. Adaptation of the OG mapping framework to the mapping of the probable locations of BP stems intersecting the NBP should be possible with the development of a forward model to describe the probability of eH detection analogous to that applied to BP detection. The information that can be gleaned from an eH anomaly about likely BP location will be greatly enhanced if the residence time in the NBP could be constrained from the magnitude of the anomaly. In terms of the binary forward model developed in Ch. 2, this information can be incorporated in the specification of the P_c^t by making grid cells at the right distance upwind more like to have emitted the reducing fluid than those closer or farther away.

Appendix A

Notation

Conventions

Typeface

x	(normal) scalar variable
\mathbf{x}	(bold) vector
\mathcal{X}	(calligraphic) set
\mathbb{R}	(blackboard bold) e.g. the set of real numbers \mathbb{R}

Time Dependence

x^t	superscripted t or τ denotes time dependence
x^τ	the set $\{x^\tau : \tau \in \tau\}$

Probabilistic Quantities

$P[A]$	probability of event A
$P[A \mid B]$	conditional probability of event A given event B
$p(x)$	probability density function of random variable x
$p(x \mid y)$	conditional probability density function of x given y
$E[x]$	expectation of x
$E[x \mid y]$	conditional expectation of x given y
H_x	entropy of x
$H_{x \mid y}(y)$	conditional entropy of x given y

Frames

$^f \mathbf{x}$	vector \mathbf{x} expressed in frame f
$_{f_2}^{f_1} R$	Rotation matrix between frames f_1 and f_2 : $_{f_2}^{f_1} \mathbf{x} = {}_{f_2}^{f_1} R {}_{f_2} \mathbf{x}$

Variable List: Occupancy Grid Mapping

Map and Cells

- μ_c random binary indicator variable $\mu_c \in \{0, 1\}$ for the state of cell c
- m_c the event that cell c is occupied: $\mu_c = 1$
- \bar{m}_c the event that cell c is empty: $\mu_c = 0$
- \mathbf{m} the map, i.e. the set $\{\mu_c: c \in \{1, \dots, C\}\}$
- C the number of cells in the map: $\mathbf{m} \in \mathbb{B}^C$
- \mathcal{S} the set of occupied cells: $\mathcal{S} \triangleq \{s: m_s, s \in \{1, \dots, C\}\}$
- $\bar{\mathcal{S}}$ the set of empty cells: $\bar{\mathcal{S}} \triangleq \{s: \bar{m}_s, s \in \{1, \dots, C\}\}$

Measurements

- z^t sensor measurement at time t
- Z^t the set of all sensor measurements up to time t
- δ^t binary measurement $\delta^t \in \{0, 1\}$
- Δ^t the set of all binary measurements
- d^t the event of a detection at time t : $\delta^t = 1$
- \bar{d}^t the event of a non-detection at time t : $\delta^t = 0$
- D^t the set of all measurements that resulted in detections
- \bar{D}^t the set of all measurements that resulted in non-detections up to time t
- n^t the number of detections up to time t : $n^t \triangleq |D^t|$
- \bar{n}^t the number of non-detections up to time t : $\bar{n}^t \triangleq |\bar{D}^t|$
- τ^t the set of measurement times that resulted in detections
- $\bar{\tau}^t$ the set of measurement times that resulted in non-detections

Selected Probabilistic Quantities

$p(\mathbf{m} \mid Z^t)$	full posterior over all maps given all measurements up to time t
$P[m_c \mid Z^t]$	marginal posterior for the occupancy of cell c given all measurements up to time t
P_c^t	(shorthand) probability that sufficient signal from cell c arrives at the sensor to trigger a detection at time t
\bar{P}_c^0	(shorthand) prior probability of occupancy for cell c : $\bar{P}_c^0 \triangleq P[m_c]$
\bar{P}_c^t	(shorthand) posterior probability of occupancy for cell c conditioned on all non-detections up to time t : $\bar{P}_c^t \triangleq P[m_c \mid \bar{D}^t]$
\tilde{P}_c^t	(shorthand) “revised prior:” the marginal posterior probability of occupancy under the assumption that the marginal posteriors are independent.
P_F^t	probability of false alarm (detection) at time t
r_c^t	odds ratio for cell c : $r_c^t \triangleq P[m_c \mid Z^t] / P[\bar{m}_c \mid Z^t]$
$\bar{\rho}_c^t$	odds ratio for cell c conditioned on exclusively non-detections: $\bar{\rho}_c^t \triangleq P[m_c \mid D^t, \bar{D}^t] / P[\bar{m}_c \mid D^t, \bar{D}^t]$
ρ_c^t	not an odds ratio but notation reflects parallel algorithmic role to $\bar{\rho}_c^t$: $\rho_c^t \triangleq P[D^t \mid \bar{D}^t, m_c] / P[D^t \mid \bar{D}^t, \bar{m}_c]$

Miscellaneous

\mathbf{g}_k^t	arbitrary grouping of measurement times that resulted in detections; k denotes group index
\mathcal{G}^t	indices k of groups \mathbf{g}_k^t

Appendix B

Exact Binary Occupancy Grid Mapping Derivations

B.1 Sums Over Maps

Sums of the form $\sum_{\mathbf{m}:m_c} \prod_{s \in \mathcal{S}} h_s \cdot \prod_{s \in \bar{\mathcal{S}}} \bar{h}_s$ appear throughout the derivations contained in this appendix. They can be simplified iteratively as follows:

$$\begin{aligned}
& \sum_{\mathbf{m}:m_c} \prod_{s \in \mathcal{S}} h_s \cdot \prod_{s \in \bar{\mathcal{S}}} \bar{h}_s \\
&= h_c \cdot \sum_{\mathbf{m}:m_c} \prod_{s \in \mathcal{S}, s \neq c} h_s \cdot \prod_{s \in \bar{\mathcal{S}}} \bar{h}_s \\
&= h_c \left(\sum_{\mathbf{m}:m_c, m_b} \prod_{s \in \mathcal{S}, s \neq c} h_s \cdot \prod_{s \in \bar{\mathcal{S}}} \bar{h}_s + \sum_{\mathbf{m}:m_c, \bar{m}_b} \prod_{s \in \mathcal{S}, s \neq c} h_s \cdot \prod_{s \in \bar{\mathcal{S}}} \bar{h}_s \right) \\
&= h_c \left(h_b \cdot \sum_{\mathbf{m}:m_c, m_b} \prod_{s \in \mathcal{S}, s \neq c, b} h_s \cdot \prod_{s \in \bar{\mathcal{S}}} \bar{h}_s + \bar{h}_b \cdot \sum_{\mathbf{m}:m_c, \bar{m}_b} \prod_{s \in \mathcal{S}, s \neq c} h_s \cdot \prod_{s \in \bar{\mathcal{S}}, s \neq b} \bar{h}_s \right) \quad (\text{B.1}) \\
&= h_c \left(h_b \cdot \sum_{\mathbf{m}:m_c, m_b} \prod_{s \in \mathcal{S}, s \neq c, b} h_s \cdot \prod_{s \in \bar{\mathcal{S}}} \bar{h}_s + \bar{h}_b \cdot \sum_{\mathbf{m}:m_c, m_b} \prod_{s \in \mathcal{S}, s \neq c, b} h_s \cdot \prod_{s \in \bar{\mathcal{S}}} \bar{h}_s \right) \\
&= h_c (h_b + \bar{h}_b) \cdot \sum_{\mathbf{m}:m_c, m_b} \prod_{s \in \mathcal{S}, s \neq c, b} h_s \cdot \prod_{s \in \bar{\mathcal{S}}} \bar{h}_s \\
&\quad \downarrow \text{(iterate over } s \in \{1, \dots, C\}, s \neq c, b) \\
&= h_c \prod_{s \neq c} (h_s + \bar{h}_s)
\end{aligned}$$

Changing the limit on the sum to all maps \mathbf{m} alters this result trivially to

$$\sum_{\mathbf{m}} \prod_{s \in \mathcal{S}} h_s \cdot \prod_{s \in \bar{\mathcal{S}}} \bar{h}_s = \prod_{s=1}^C (h_s + \bar{h}_s) . \quad (\text{B.2})$$

B.2 Derivation: Binary Inverse Model

To derive (2.14) in the main text, begin by applying Bayes Rule to $P[d^t | \mathbf{m}]$ and $P[\bar{d}^t | \mathbf{m}]$, and then marginalize:

$$P[m_c | d^t] = \frac{\sum_{\mathbf{m}:m_c} P[d^t | \mathbf{m}] p(\mathbf{m})}{\sum_{\mathbf{m}} P[d^t | \mathbf{m}] p(\mathbf{m})} \quad (\text{B.3a})$$

$$P[m_c | \bar{d}^t] = \frac{\sum_{\mathbf{m}:m_c} P[\bar{d}^t | \mathbf{m}] p(\mathbf{m})}{\sum_{\mathbf{m}} P[\bar{d}^t | \mathbf{m}] p(\mathbf{m})} \quad (\text{B.3b})$$

where $p(\mathbf{m})$ denotes the prior on the map. The state of each grid cell is assumed to be independent so that $p(\mathbf{m}) = \prod_{s \in \mathcal{S}} \bar{P}_s^0 \cdot \prod_{s \in \bar{\mathcal{S}}} (1 - \bar{P}_s^0)$ where \bar{P}_c^0 denotes the single-cell prior $P[m_c]$.

To proceed we substitute (2.11) into the numerator of (B.3):

$$\begin{aligned} & \sum_{\mathbf{m}:m_c} P[d^t | \mathbf{m}] p(\mathbf{m}) \\ &= \sum_{\mathbf{m}:m_c} (1 - (1 - P_F^t) \prod_{s \in \mathcal{S}} (1 - P_s^t)) \prod_{s \in \mathcal{S}} \bar{P}_s^t \prod_{s \in \bar{\mathcal{S}}} (1 - \bar{P}_s^t) \\ &= \bar{P}_c^0 - (1 - P_F^t) \sum_{\mathbf{m}:m_c} \prod_{s \in \mathcal{S}} (1 - P_s^t) \bar{P}_s^0 \prod_{s \in \bar{\mathcal{S}}} (1 - \bar{P}_s^0) \\ &= \bar{P}_c^0 - (1 - P_F^t)(1 - P_c^t) \bar{P}_c^0 \prod_{s \neq c} ((1 - P_s^t) \bar{P}_s^0 + 1 - \bar{P}_s^0) \\ &= \bar{P}_c^0 \left(1 - (1 - P_F^t)(1 - P_c^t) \prod_{s \neq c} (1 - P_s^t \bar{P}_s^0) \right) \end{aligned} \quad (\text{B.4})$$

where the third line follows from the result for generic sums of this form (App. B.1).

Upon noting that $\sum_{\mathbf{m}:m_c} P[\bar{d}^t | \mathbf{m}] p(\mathbf{m}) = \sum_{\mathbf{m}:m_c} (1 - P[d^t | \mathbf{m}]) p(\mathbf{m})$, the remaining sums appearing in (B.3) become trivial modifications of (B.4). Substituting

the resulting expressions into (B.3) leads, after some straightforward simplification, to the complete inverse sensor model given as (2.14) in the text.

B.3 Proof: $p(\mathbf{m} \mid \bar{D}^t) = \prod_{s \in \mathcal{S}} \bar{P}_s^t \cdot \prod_{s \in \bar{\mathcal{S}}} (1 - \bar{P}_s^t)$

It is shown that the conditional independence of cells holds if conditioned only on nondetections:

$$p(\mathbf{m} \mid \bar{D}^t) = \prod_{s \in \mathcal{S}} \bar{P}_s^t \cdot \prod_{s \in \bar{\mathcal{S}}} (1 - \bar{P}_s^t) , \quad (\text{B.5})$$

and furthermore that a binary forward model of the form (2.11) is both the necessary and sufficient condition.

Proof. The proof of sufficiency is by induction. Let $Q(t)$ be the proposition that

$$p(\mathbf{m} \mid \bar{D}^t) = \prod_{s \in \mathcal{S}} \bar{P}_s^t \cdot \prod_{s \in \bar{\mathcal{S}}} (1 - \bar{P}_s^t)$$

and

$$\bar{P}_c^{t+1} = \frac{(1 - P_c^{t+1}) \bar{P}_c^t}{1 - P_c^{t+1} \bar{P}_c^t} .$$

In the base case, the first portion of $Q(0)$ is true because $p(\mathbf{m} \mid \bar{D}^0) = p(\mathbf{m})$ by definition and the map nodes are independent so that

$$p(\mathbf{m}) = \prod_{s \in \mathcal{S}} \bar{P}_s^0 \cdot \prod_{s \in \bar{\mathcal{S}}} (1 - \bar{P}_s^0) .$$

The second portion of $Q(0)$ is also true, by our previous result for the inverse sensor model (2.14b).

For $t \geq 0$, applying Bayes Rule to $p(\mathbf{m} \mid \bar{D}^{t+1})$ yields

$$\begin{aligned} p(\mathbf{m} \mid \bar{D}^{t+1}) &= \frac{p(\bar{d}^{t+1} \mid \bar{D}^t, \mathbf{m}) p(\mathbf{m} \mid \bar{D}^t)}{\sum_{\mathbf{m}} p(\bar{d}^{t+1} \mid \bar{D}^t, \mathbf{m}) p(\mathbf{m} \mid \bar{D}^t)} \\ &= \frac{p(\bar{d}^{t+1} \mid \mathbf{m}) p(\mathbf{m} \mid \bar{D}^t)}{\sum_{\mathbf{m}} p(\bar{d}^{t+1} \mid \mathbf{m}) p(\mathbf{m} \mid \bar{D}^t)} \end{aligned} \quad (\text{B.6})$$

where the second line follows from the “static world” assumption.

By induction the term $p(\mathbf{m} \mid \bar{D}^t)$ can be expressed as a product over all cells so that the sum appearing in the denominator of (B.6) is of the same form as (B.4). Thus $\sum_{\mathbf{m}} p(\bar{d}^{t+1} \mid \mathbf{m}) p(\mathbf{m} \mid \bar{D}^t) = (1 - P_F^{t+1}) \prod_s (1 - P_s^{t+1} \bar{P}_s^t)$.

With this substitution and our expression (2.11) for the forward measurement model $P[\bar{d}^t \mid \mathbf{m}]$, (B.6) becomes

$$\begin{aligned}
p(\mathbf{m} \mid \bar{D}^{t+1}) &= \frac{(1 - P_F^{t+1}) \prod_{s \in \mathcal{S}} (1 - P_s^{t+1})}{(1 - P_F^{t+1}) \prod_s (1 - P_s^{t+1} \bar{P}_s^t)} p(\mathbf{m} \mid \bar{D}^t) \\
&= \prod_{s \in \mathcal{S}} \frac{(1 - P_s^{t+1}) \bar{P}_s^t}{1 - P_s^{t+1} \bar{P}_s^t} \cdot \prod_{s \in \bar{\mathcal{S}}} \frac{1 - \bar{P}_s^t}{1 - P_s^{t+1} \bar{P}_s^t} \\
&= \prod_{s \in \mathcal{S}} \frac{(1 - P_s^{t+1}) \bar{P}_s^t}{1 - P_s^{t+1} \bar{P}_s^t} \cdot \prod_{s \in \bar{\mathcal{S}}} \left(1 - \frac{(1 - P_s^{t+1}) \bar{P}_s^t}{1 - P_s^{t+1} \bar{P}_s^t} \right) \\
&= \prod_{s \in \mathcal{S}} \bar{P}_s^{t+1} \cdot \prod_{s \in \bar{\mathcal{S}}} (1 - \bar{P}_s^{t+1})
\end{aligned} \tag{B.7}$$

which completes the proof. \square

That the form of $P[\bar{d}^t \mid \mathbf{m}]$ given by (2.11) is also a necessary condition for (B.5) to hold is shown by direct proof.

Proof. Begin by applying Bayes Rule to $p(\bar{D}^t \mid \mathbf{m})$ and then factor $p(\mathbf{m} \mid \bar{D}^t)$ using the independence property proved above:

$$\begin{aligned}
p(\bar{D}^t \mid \mathbf{m}) &= \frac{p(\mathbf{m} \mid \bar{D}^t) p(\bar{D}^t)}{p(\mathbf{m})} \\
&= \frac{p(\bar{D}^t)}{p(\mathbf{m})} \cdot \prod_{s \in \mathcal{S}} P[m_s \mid \bar{D}^t] \cdot \prod_{s \in \bar{\mathcal{S}}} P[\bar{m}_s \mid \bar{D}^t] \\
&= p(\bar{D}^t) \cdot \prod_{s \in \mathcal{S}} \frac{P[m_s \mid \bar{D}^t] P[\bar{m}_s]}{P[\bar{m}_s \mid \bar{D}^t] P[m_s]} \cdot \prod_{s=1}^C \frac{P[\bar{m}_s \mid \bar{D}^t]}{P[\bar{m}_s]}.
\end{aligned} \tag{B.8}$$

By (B.5) the latter product in the last line of (B.8) is

$$\begin{aligned}
\prod_{s=1}^C \frac{P[\bar{m}_s \mid \bar{D}^t]}{P[\bar{m}_s]} &= \frac{P[\mathcal{S} = \emptyset \mid \bar{D}^t]}{P[\mathcal{S} = \emptyset]} \\
&= \frac{P[\bar{D}^t \mid \mathcal{S} = \emptyset]}{p(\bar{D}^t)},
\end{aligned} \tag{B.9}$$

where the second line follows by Bayes Rule.

Similarly,

$$\begin{aligned} P[m_c | \bar{D}^t] \cdot \prod_{s \neq c} P[m_s | \bar{D}^t] &= P[\mathcal{S} = c | \bar{D}^t] \\ &= \frac{p(\bar{D}^t | \mathcal{S} = c) P[\mathcal{S} = c]}{p(\bar{D}^t)}. \end{aligned} \quad (\text{B.10})$$

Rearranging terms and using the independence of the prior,

$$\frac{P[m_c | \bar{D}^t]}{P[m_c]} = \frac{p(\bar{D}^t | \mathcal{S} = c)}{p(\bar{D}^t)} \cdot \prod_{s \neq c} \frac{P[\bar{m}_s]}{P[m_s | \bar{D}^t]}. \quad (\text{B.11})$$

With this result the argument of the first product in the last line of (B.8) becomes

$$\begin{aligned} \frac{P[m_c | \bar{D}^t] P[\bar{m}_c]}{P[\bar{m}_c | \bar{D}^t] P[m_c]} &= \frac{p(\bar{D}^t | \mathcal{S} = c)}{p(\bar{D}^t)} \cdot \prod_{s=1}^C \frac{P[\bar{m}_s]}{P[\bar{m}_c | \bar{D}^t]} \\ &= \frac{p(\bar{D}^t | \mathcal{S} = c)}{p(\bar{D}^t | \mathcal{S} = \emptyset)} \end{aligned} \quad (\text{B.12})$$

where the last line follows upon substitution of the result from (B.9).

The probabilities in results (B.9) and (B.12) can be evaluated in terms of the definitions for P_c^t and P_F^t by taking advantage of the static world assumption:

$$\begin{aligned} p(\bar{D}^t | \mathcal{S} = \emptyset) &= \prod_{\tau \in \bar{\tau}^t} P[\bar{d}^\tau | \mathcal{S} = \emptyset] \\ &= \prod_{\tau \in \bar{\tau}^t} (1 - P_F^\tau) \end{aligned} \quad (\text{B.13})$$

and

$$\begin{aligned} p(\bar{D}^t | \mathcal{S} = c) &= \prod_{\tau \in \bar{\tau}^t} P[\bar{d}^\tau | \mathcal{S} = c] \\ &= \prod_{\tau \in \bar{\tau}^t} (1 - P_F^\tau)(1 - P_c^\tau). \end{aligned} \quad (\text{B.14})$$

Substituting these results into (B.9) and (B.12) yields

$$\prod_{s=1}^C \frac{P[m_s | \bar{D}^t]}{P[\bar{m}_s]} = \frac{1}{p(\bar{D}^t)} \prod_{\tau \in \bar{\tau}^t} (1 - P_F^\tau) \quad (\text{B.15a})$$

$$\prod_{s \in \mathcal{S}} \frac{P[m_s | \bar{D}^t] P[\bar{m}_s]}{P[\bar{m}_s | \bar{D}^t] P[m_s]} = \prod_{s \in \mathcal{S}} \prod_{\tau \in \bar{\tau}^t} (1 - P_c^\tau) . \quad (\text{B.15b})$$

Finally, substituting (B.15a) and (B.15b) into (B.8) gives

$$\begin{aligned} p(\bar{D}^t | \mathbf{m}) &= \prod_{\tau \in \bar{\tau}^t} (1 - P_F^\tau) \cdot \prod_{s \in \mathcal{S}} \prod_{\tau \in \bar{\tau}^t} (1 - P_s^\tau) \\ &= \prod_{\tau \in \bar{\tau}^t} \left((1 - P_F^\tau) \cdot \prod_{s \in \mathcal{S}} (1 - P_s^\tau) \right) \end{aligned} \quad (\text{B.16})$$

This result must be also be true for $\bar{\tau}^t = t$, thus $P[\bar{d}^t | \mathbf{m}]$ as given by (2.13) is not only sufficient, but also necessary for (B.8) to hold. \square

B.4 Proof: $P[\bar{d}^t | \bar{D}^{t-1}, m_c] \neq P[\bar{d}^t | m_c]$

The conditional independence assumption required by the standard OG mapping algorithm is $P[z^t | Z^{t-1}, m_c] = P[z^t | m_c]$. This assumption remains false even when all measurements consist solely of non-detections ($Z^t = \bar{D}^t$). This can be shown by exploiting (2.17), proven above, to derive exact expressions for $P[\bar{d}^t | m_c]$ and $P[\bar{d}^t | \bar{D}^{t-1}, m_c]$.

Beginning with $P[\bar{d}^t | m_c]$,

$$\begin{aligned} P[\bar{d}^t | m_c] &= \frac{1}{P[m_c]} \sum_{\mathbf{m}: m_c} P[\bar{d}^t | \mathbf{m}] p(\mathbf{m}) \\ &= (1 - P_c^t) \sum_{\mathbf{m}: m_c} \prod_{s \in \mathcal{S}, s \neq c} (1 - P_s^t) P[m_s] \cdot \prod_{s \in \bar{\mathcal{S}}} P[\bar{m}_s] \\ &= (1 - P_c^t) \prod_{s \neq c} \left((1 - P_s^t) P[m_s] + 1 - P[m_s] \right) \\ &= (1 - P_c^t) \prod_{s \neq c} (1 - P_s^t P[m_s]) . \end{aligned} \quad (\text{B.17})$$

Similarly, for $P [\bar{d}^t \mid \bar{D}^t, m_c]$,

$$\begin{aligned}
P [\bar{d}^t \mid \bar{D}^{t-1}, m_c] &= \frac{1}{P[m_c]} \sum_{\mathbf{m}: m_c} P [\bar{d}^t \mid \bar{D}^{t-1}, \mathbf{m}] p(\mathbf{m} \mid \bar{D}^{t-1}) \\
&= \frac{1}{P[m_c]} \sum_{\mathbf{m}: m_c} P [\bar{d}^t \mid \mathbf{m}] p(\mathbf{m} \mid \bar{D}^{t-1}) \\
&= (1 - P_c^t) \sum_{\mathbf{m}: m_c} \prod_{s \in \mathcal{S}, s \neq c} (1 - P_s^t) P [m_s \mid \bar{D}^{t-1}] \cdot \prod_{s \in \bar{\mathcal{S}}} P [\bar{m}_s \mid \bar{D}^{t-1}] \\
&= (1 - P_c^t) \prod_{s \neq c} (1 - P_s^t P [m_s \mid \bar{D}^{t-1}]) .
\end{aligned} \tag{B.18}$$

The second line above follows from the static world assumption. The third line exploits (2.17) to factor $p(\mathbf{m} \mid \bar{D}^{t-1})$.

It may be shown readily that

$$P [m_c \mid \bar{D}^{t-1}] \leq P [m_c] \tag{B.19}$$

where equality holds only for the trivial conditions $P_c^\tau = 0 \forall \tau$ or $P [m_c] \in \{0, 1\}$. Consequently,

$$(1 - P_c^t) \prod_{s \neq c} (1 - P_s^t P [m_s \mid \bar{D}^{t-1}]) > (1 - P_c^t) \prod_{s \neq c} (1 - P_s^t P [m_s]) \tag{B.20}$$

where the strict inequality holds unless $t = 1$, $P_c^t = 1$, or at least one of the other trivial conditions is met for each $s \neq c$. Thus the standard OG mapping algorithm assumption remains invalid even when applied exclusively to non-detections.

The inequality (B.19) is most readily derived from the expression for the odds ratio conditioned solely on non-detections (2.20) from which it follows that

$$P [m_c \mid \bar{D}^{t-1}] = \frac{\prod_{\tau=1}^{t-1} (1 - P_c^\tau)}{1 - \left(1 - \prod_{\tau=1}^{t-1} (1 - P_c^\tau)\right) P [m_c]} P [m_c] . \tag{B.21}$$

To show (B.19) it is therefore sufficient to show

$$\prod_{\tau=1}^{t-1} (1 - P_c^\tau) \leq 1 - \left(1 - \prod_{\tau=1}^{t-1} (1 - P_c^\tau) \right) P[m_c] . \quad (\text{B.22})$$

Rearranging terms,

$$\prod_{\tau=1}^{t-1} (1 - P_c^\tau) (1 - P[m_c]) \leq (1 - P[m_c]) \quad (\text{B.23})$$

which is obviously true. Equality holds iff $P_c^\tau = 0 \forall \tau$ or $P[m_c] = 1$. From (B.21), $P[m_c \mid \bar{D}^{t-1}] = P[m_c]$ also holds under the condition $P[m_c] = 0$.

B.5 Derivation: Exact Solution

This section derives the expressions for the exact posterior odds ratios given by (2.20) and (2.21). The posterior odds ratio conditioned on exclusively non-detections (2.20) is readily attained by application of (B.12)–(B.14):

$$\begin{aligned} \bar{\rho}_c^t &\triangleq \frac{P[m_c \mid \bar{D}^t]}{P[\bar{m}_c \mid \bar{D}^t]} \\ &= \frac{p(\bar{D}^t \mid \mathcal{S} = c)}{p(\bar{D}^t \mid \mathcal{S} = \emptyset)} \cdot \frac{P[m_c]}{P[\bar{m}_c]} \\ &= \prod_{\tau \in \tau^t} (1 - P_c^\tau) \cdot \frac{\bar{P}_c^0}{1 - \bar{P}_c^0} . \end{aligned} \quad (\text{B.24})$$

Expressed in recursive form,

$$\bar{\rho}_c^t = (1 - P_c^t) \bar{\rho}_c^{t-1} \quad (\text{B.25})$$

as given in the text.

To derive the expression for the complete posterior odds ratio (2.21) begin by applying Bayes Rule to the marginal posterior, making use of the static world assumption to split up $p(D^t, \bar{D}^t \mid \mathbf{m})$, and then reapply Bayes rule to just the term

containing \bar{D}^t :

$$\begin{aligned}
P[m_c | D^t, \bar{D}^t] &= \sum_{\mathbf{m}:m_c} p(\mathbf{m} | D^t, \bar{D}^t) \\
&= \sum_{\mathbf{m}:m_c} \frac{p(D^t, \bar{D}^t | \mathbf{m}) p(\mathbf{m})}{p(D^t, \bar{D}^t)} \\
&= \frac{1}{p(D^t, \bar{D}^t)} \sum_{\mathbf{m}:m_c} p(D^t | \mathbf{m}) p(\bar{D}^t | \mathbf{m}) p(\mathbf{m}) \\
&= \frac{p(\bar{D}^t)}{p(D^t, \bar{D}^t)} \sum_{\mathbf{m}:m_c} p(D^t | \mathbf{m}) p(\mathbf{m} | \bar{D}^t) .
\end{aligned}$$

Upon substituting in our expression for the forward model (2.11),

$$\begin{aligned}
P[m_c | D^t, \bar{D}^t] &= \frac{p(\bar{D}^t)}{p(D^t, \bar{D}^t)} \\
&\quad \times \sum_{\mathbf{m}:m_c} \prod_{\tau_i \in \tau^t} \left(1 - (1 - P_F^{\tau_i}) \prod_{s \in \mathcal{S}} (1 - P_s^{\tau_i}) \right) p(\mathbf{m} | \bar{D}^t) . \quad (\text{B.26})
\end{aligned}$$

Note that the dependence on the prior $p(\mathbf{m})$ has been replaced by a dependence on $p(\mathbf{m} | \bar{D}^t)$ which nonetheless retains the prior's multiplicative structure by virtue of the conditional independence proved in App. B.3.

To proceed we expand the product over D^t as a sum:

$$\begin{aligned}
&\prod_{\tau_i \in \tau^t} \left(1 - (1 - P_F^{\tau_i}) \prod_{s \in \mathcal{S}} (1 - P_s^{\tau_i}) \right) \\
&= 1 + \sum_{q=1}^{n^t} (-1)^q \sum_{\substack{\boldsymbol{\eta} \subseteq \tau^t \\ \boldsymbol{\eta} \in \mathbb{N}^q}} \prod_{i=1}^q \left((1 - P_F^{\eta_i}) \prod_{s \in \mathcal{S}} (1 - P_s^{\eta_i}) \right) .
\end{aligned}$$

This expression allows the order of summation in (B.26) to be changed:

$$\begin{aligned}
P[m_c | D^t, \bar{D}^t] &= \frac{p(\bar{D}^t)}{p(D^t, \bar{D}^t)} \left[\bar{P}_c^t \right. \\
&\quad \left. + \sum_{q=1}^{n^t} (-1)^q \sum_{\substack{\boldsymbol{\eta} \subseteq \tau^t \\ \boldsymbol{\eta} \in \mathbb{N}^q}} \sum_{\mathbf{m}:m_c} \prod_{i=1}^q \left((1 - P_F^{\eta_i}) \prod_{s \in \mathcal{S}} (1 - P_s^{\eta_i}) \right) p(\mathbf{m} | \bar{D}^t) \right] . \quad (\text{B.27})
\end{aligned}$$

To arrive at the above, we have used the conditional independence of $p(\mathbf{m} \mid \bar{D}^t)$ to get $\sum_{\mathbf{m}:m_c} p(\mathbf{m} \mid \bar{D}^t) = \bar{P}_c^t$.

Working now with just the sum over $\mathbf{m} : m_c$,

$$\begin{aligned}
& \sum_{\mathbf{m}:m_c} \prod_{i=1}^q \left((1 - P_F^{\eta_i}) \prod_{s \in \mathcal{S}} (1 - P_s^{\eta_i}) \right) p(\mathbf{m} \mid \bar{D}^t) \\
&= \prod_{i=1}^q (1 - P_F^{\eta_i}) \\
&\quad \times \sum_{\mathbf{m}:m_c} \prod_{s \in \mathcal{S}} \left(\prod_{i=1}^q (1 - P_s^{\eta_i}) \bar{P}_s^t \right) \prod_{s \in \bar{\mathcal{S}}} (1 - \bar{P}_s^t) \\
&= \bar{P}_c^t \prod_{i=1}^q (1 - P_c^{\eta_i}) \cdot \prod_{i=1}^q (1 - P_F^{\eta_i}) \\
&\quad \times \sum_{\mathbf{m}:m_c} \prod_{\substack{s \in \mathcal{S} \\ s \neq c}} \left(\prod_{i=1}^q (1 - P_s^{\eta_i}) \bar{P}_s^t \right) \prod_{s \in \bar{\mathcal{S}}} (1 - \bar{P}_s^t) .
\end{aligned} \tag{B.28}$$

The remaining sum in (B.28) can be converted into a product over the much smaller space $s \neq c$ in the usual way:

$$\begin{aligned}
& \prod_{i=1}^q (1 - P_F^{\eta_i}) \sum_{\substack{\mathbf{m}:m_c \\ s \in \mathcal{S} \\ s \neq c}} \prod_{i=1}^q \left(\prod_{i=1}^q (1 - P_s^{\eta_i}) \bar{P}_s^t \right) \prod_{s \in \bar{\mathcal{S}}} (1 - \bar{P}_s^t) \\
&= \prod_{i=1}^q (1 - P_F^{\eta_i}) \prod_{s \neq c} \left(\bar{P}_s^t \prod_{i=1}^q (1 - P_s^{\eta_i}) + (1 - \bar{P}_s^t) \right) \\
&= \prod_{i=1}^q (1 - P_F^{\eta_i}) \prod_{s \neq c} \left(1 - \left(1 - \prod_{i=1}^q (1 - P_s^{\eta_i}) \right) \bar{P}_s^t \right) \\
&= f_c^q(\boldsymbol{\eta}) .
\end{aligned}$$

Substituting this result into (B.28) and the resulting expression into (B.27) yields the

posterior probability of occupancy:

$$\begin{aligned} \text{P} [m_c \mid D^t, \bar{D}^t] \\ = \frac{\text{p}(\bar{D}^t)}{\text{p}(D^t, \bar{D}^t)} \left[1 + \sum_{q=1}^{n^t} (-1)^q \sum_{\substack{\boldsymbol{\eta} \subseteq \boldsymbol{\tau}^t \\ \boldsymbol{\eta} \in \mathbb{N}^q}} \prod_{i=1}^q (1 - P_c^{\eta_i}) f_c^q(\boldsymbol{\eta}) \right] \bar{P}_c^t . \quad (\text{B.29}) \end{aligned}$$

A nearly identical procedure carried out for the posterior probability of emptiness leads to:

$$\begin{aligned} \text{P} [\bar{m}_c \mid D^t, \bar{D}^t] \\ = \frac{\text{p}(\bar{D}^t)}{\text{p}(D^t, \bar{D}^t)} \left[1 + \sum_{q=1}^{n^t} (-1)^q \sum_{\substack{\boldsymbol{\eta} \subseteq \boldsymbol{\tau}^t \\ \boldsymbol{\eta} \in \mathbb{N}^q}} f_c^q(\boldsymbol{\eta}) \right] (1 - \bar{P}_c^t) . \quad (\text{B.30}) \end{aligned}$$

The unknown leading coefficient is identical for both occupancy and emptiness and cancels upon constructing the odds ratio.

Appendix C

Exact OG mapping algorithm with unconnected observation regions

This implementation of Algorithm 1 exploits the lack of dependence between detections that observe portions of the map not related to one another either by direct overlap or through overlap with shared neighbors.

Let $\mathcal{R}^t = \{c: P_c^t > 0\}$ denote the portion of the map observed by a measurement at time t , and let \mathcal{R}_*^t denote the union of all connected \mathcal{R}^τ for $\tau \in \tau^{t-1}$ such that at least one $\mathcal{R}^{t\tau} \subset \mathcal{R}_*^t$ overlaps with \mathcal{R}^t . Note that \mathcal{R}_*^t is exclusive of any cells belonging to \mathcal{R}^t alone and may consist of disjoint subsets of the \mathcal{R}^τ . The definition of \mathcal{R}_*^t is illustrated schematically in Fig. C-1 and Algorithm 10 implements a recursive procedure in pseudocode for finding \mathcal{R}_*^t from a seed region.

If a detection was registered at time t , then the posteriors of cells within $\mathcal{R}^t \cup \mathcal{R}_*^t$ are dependent. If on the other hand a non-detection occurred, then all disjoint sets of connected regions remain independent from one another, regardless of any shared history of observations that resulted in non-detections. Indeed, it can be shown that all terms in (2.21) containing P_s^τ for all s not in the same connected region as c factor from the numerator and denominator and cancel. Figure C-2 illustrates the independent update of connected regions using the sonar range finder model of Fig. 2-3.

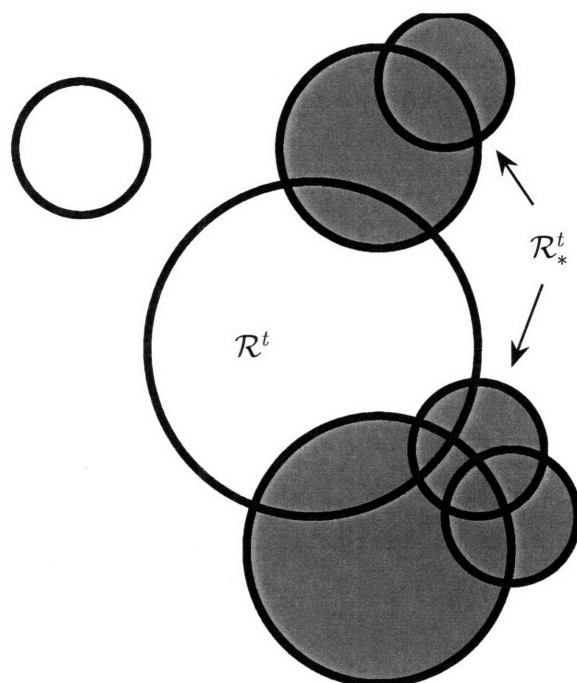


Figure C-1: Schematic definitions of \mathcal{R}^t , the portion of the map observed at time t , and the previous detection regions that comprise \mathcal{R}_*^t , indicated in gray.

Algorithm 9 An exact algorithm for the odds ratio. The algorithm requires the subroutine **find_connected** (10) to determine the portions of the map that require updating. Unless otherwise noted, all subscripts c imply the operation is for all $c \in \{1, \dots, C\}$.

Require:

$\rho_c^{t-1}, \bar{\rho}_c^{t-1}$
 P_c^t
 $P_c^{\tau_i}, \forall \tau_i \in \tau^{t-1}$
 $\mathcal{R}^t \triangleq \{c : P_c^t > 0\}$ {current measurement region}
 $\mathcal{R}^{\tau_i}, \forall \tau_i \in \tau^{t-1}$ {all previous detection regions}

```

 $r_c^t = \text{og\_exact}\{$ 
  1: if  $d^t$  (detection) then
  2:    $\bar{\rho}_c^t = \bar{\rho}_c^{t-1}$ 
  3:    $\bar{P}_c^t = \bar{\rho}_c^t / (1 + \bar{\rho}_c^t)$ 
  4:    $\mathcal{R}_*^t = \text{find\_connected}(\mathcal{R}^t)$ 
  5:   for all  $c \in \mathcal{R}_*^t \cup \mathcal{R}^t$  do
  6:     
$$\rho_c^t = \frac{1 + \sum_{q=1}^{n^t} (-1)^q \sum_{\eta \subseteq \tau^t, \eta \in \mathcal{N}^q} \prod_{i=1}^q (1 - P_c^{\eta_i}) \prod_{i=1}^{|\eta|} (1 - P_F^{\eta_i}) \prod_{s \neq c, s \in \mathcal{R}_*^t \cup \mathcal{R}^t} \left( 1 - \left( 1 - \prod_{i=1}^{|\eta|} (1 - P_s^{\eta_i}) \right) \bar{P}_s^t \right)}{1 + \sum_{q=1}^{n^t} (-1)^q \sum_{\eta \subseteq \tau^t, \eta \in \mathcal{N}^q} \prod_{i=1}^{|\eta|} (1 - P_F^{\eta_i}) \prod_{s \neq c, s \in \mathcal{R}_*^t \cup \mathcal{R}^t} \left( 1 - \left( 1 - \prod_{i=1}^{|\eta|} (1 - P_s^{\eta_i}) \right) \bar{P}_s^t \right)}$$

  7:   end for
  8:    $\rho_c^t = \rho_c^{t-1}, \forall c \notin \mathcal{R}_*^t$ 
  9: else if  $\bar{d}^t$  (nondetection) then
  10:   $\bar{\rho}_c^t = (1 - P_c^t) \bar{\rho}_c^{t-1}, \forall c \in \mathcal{R}^t$ 
  11:   $\bar{\rho}_c^t = \bar{\rho}_c^{t-1}, \forall c \notin \mathcal{R}^t$ 
  12:   $\mathcal{R}_*^t = \text{find\_connected}(\mathcal{R}^t)$ 
  13:   $\rho_c^t = \rho_c^{t-1}, \forall c \notin \mathcal{R}_*^t$  {The  $\bar{\rho}_c^t$  outside of  $\mathcal{R}_*^t$  are unchanged}
    {Each disjoint subset of connected regions in  $\mathcal{R}_*^t$  can be processed independently}
  14:   $\tau_{**} = \{\tau : \tau < t, \mathcal{R}^\tau \cap \mathcal{R}^t \neq \emptyset\}$ 
  15:  while  $\exists \tau \in \tau_{**}$  such that  $\mathcal{R}^\tau \cap \mathcal{R}^t \neq \emptyset$  do
  16:     $\mathcal{R}_*^t = \text{find\_connected}(\mathcal{R}^\tau)$ 
  17:    for all  $c \in \mathcal{R}_*^t$  do
  18:      
$$\rho_c^t = \frac{1 + \sum_{q=1}^{n^t} (-1)^q \sum_{\eta \subseteq \tau^t, \eta \in \mathcal{N}^q} \prod_{i=1}^q (1 - P_c^{\eta_i}) \prod_{i=1}^{|\eta|} (1 - P_F^{\eta_i}) \prod_{s \neq c, s \in \mathcal{R}_*^t} \left( 1 - \left( 1 - \prod_{i=1}^{|\eta|} (1 - P_s^{\eta_i}) \right) \bar{P}_s^t \right)}{1 + \sum_{q=1}^{n^t} (-1)^q \sum_{\eta \subseteq \tau^t, \eta \in \mathcal{N}^q} \prod_{i=1}^{|\eta|} (1 - P_F^{\eta_i}) \prod_{s \neq c, s \in \mathcal{R}_*^t} \left( 1 - \left( 1 - \prod_{i=1}^{|\eta|} (1 - P_s^{\eta_i}) \right) \bar{P}_s^t \right)}$$

  19:    end for
    {Remove all processed detection times from those remaining to be processed.}
  20:   $\tau_{**} = \tau_{**} \setminus \{\tau : \mathcal{R}^\tau \cap \mathcal{R}_*^t \neq \emptyset\}$ 
  21: end while
  22: end if
  23:  $r_c^t = \rho_c^t \cdot \bar{\rho}_c^t$ 
 $\}$ 

```

Algorithm 10 Subroutine for computing a connected set of detection regions given a seed region. The result returned is exclusive of the seed region. Required for Algorithm 1.

Require:

$\mathcal{R}^t \triangleq \{c : P_c^t > 0\}$ {current measurement region}
 $\mathcal{R}^{\tau_i}, \forall \tau_i \in \tau^{t-1}$ {all previous detection regions}

$\mathcal{R}_*^t = \text{find_connected}(\mathcal{R}_*^t) \{$
 1: $\mathcal{R}_*^- = \emptyset$
 2: $\tau_* = \emptyset$
 3: **while** $\mathcal{R}_*^t \cap \mathcal{R}_*^- \neq \mathcal{R}_*^t$ **do**
 4: $\mathcal{R}_*^- = \mathcal{R}_*^t$
 5: $\tau_* = \{\tau : \tau < t, \mathcal{R}^\tau \cap \mathcal{R}_*^- \neq \emptyset\}$
 6: $\mathcal{R}_*^t = \bigcup_{\tau \in \tau_*} \mathcal{R}^\tau$
 7: **end while**
 $\}$

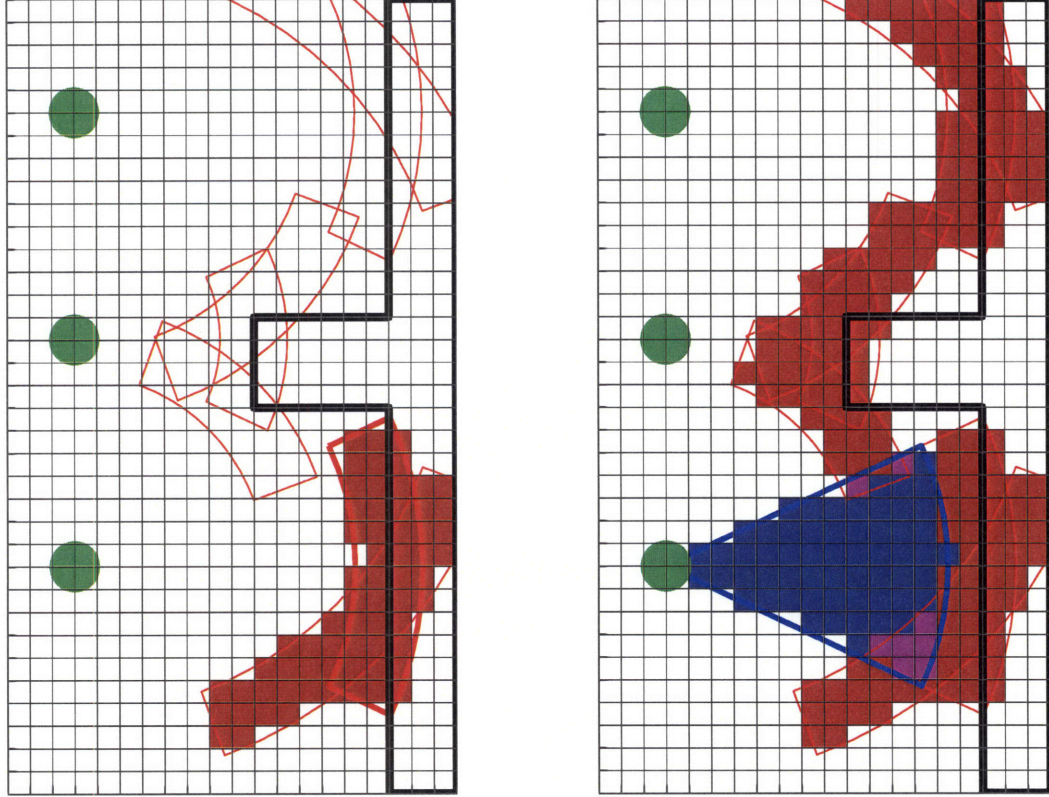


Figure C-2: Map cells updated by Algorithm 1 following a detection (left) and a non-detection (right) using the sonar model of Fig. 2-3. At each successive robot position (green circles), the robot acquired three range measurements at headings of 45° , 90° , and 135° . The detection regions for all measurements are outlined in red. Red-filled regions indicate cells for which ρ_c^t was updated following the third 90° measurement (thick lines). Similarly, blue-filled cells indicate cells for which $\bar{\rho}_c^t$ was updated, and purple-filled regions indicate cells for which both ρ_c^t and $\bar{\rho}_c^t$ required updating. The two disjoint sets of detection regions following the non-detection would be updated independently by Algorithm 1.

Appendix D

IP Algorithm Derivations

D.1 Recursive IP

To show (3.19a) in the main text, begin by applying Bayes Rule to the definition of r_c^t , followed by the total probability theorem and then the static world assumption:

$$\begin{aligned}
 r_c^t &= \frac{P[m_c \mid D^t, \bar{D}^t]}{P[\bar{m}_c \mid D^t, \bar{D}^t]} \\
 &= \frac{P[D^t, \bar{D}^t \mid m_c] P[m_c]}{P[D^t, \bar{D}^t \mid \bar{m}_c] P[\bar{m}_c]} \\
 &= \frac{\sum_{\mathbf{m}: m_c} P[D^t, \bar{D}^t \mid \mathbf{m}] p(\mathbf{m})}{\sum_{\mathbf{m}: \bar{m}_c} P[D^t, \bar{D}^t \mid \mathbf{m}] p(\mathbf{m})} \\
 &= \frac{\sum_{\mathbf{m}: m_c} \prod_{\tau=\tau_n+1}^t P[\bar{d}^\tau \mid \mathbf{m}] P[D^{\tau_n}, \bar{D}^{\tau_n} \mid \mathbf{m}] p(\mathbf{m})}{\sum_{\mathbf{m}: \bar{m}_c} \prod_{\tau=\tau_n+1}^t P[\bar{d}^\tau \mid \mathbf{m}] P[D^{\tau_n}, \bar{D}^{\tau_n} \mid \mathbf{m}] p(\mathbf{m})}.
 \end{aligned} \tag{D.1}$$

Upon substitution of the forward model for a non-detection (2.13) and application

of Bayes rule to $P [D^{\tau_n}, \bar{D}^{\tau_n} \mid \mathbf{m}]$,

$$r_c^t = \frac{\prod_{\tau=\tau_n+1}^t (1 - P_F^\tau) \sum_{\mathbf{m}:\bar{\mathbf{m}}_c} \prod_{s \in \mathcal{S}} \left(\prod_{\tau=\tau_n+1}^t (1 - P_s^\tau) \right) \cdot P [\mathbf{m} \mid D^{\tau_n}, \bar{D}^{\tau_n}]}{\prod_{\tau=\tau_n+1}^t (1 - P_F^\tau) \sum_{\mathbf{m}:\bar{\mathbf{m}}_c} \prod_{s \in \mathcal{S}} \left(\prod_{\tau=\tau_n+1}^t (1 - P_s^\tau) \right) \cdot P [\mathbf{m} \mid D^{\tau_n}, \bar{D}^{\tau_n}]} . \quad (\text{D.2})$$

Now apply the assumption $p(\mathbf{m} \mid \Delta^{\tau_n}) = \prod_{s=1}^C p(\mu_c \mid \Delta^{\tau_n})$, so that

$$P [\mathbf{m} \mid \boldsymbol{\tau}^{\tau_n}, \bar{\boldsymbol{\tau}}^{\tau_n}] = \prod_{s \in \mathcal{S}} \tilde{P}_s^{\tau_n} \prod_{s \in \bar{\mathcal{S}}} (1 - \tilde{P}_s^{\tau_n})$$

and then simplify the resulting sums in the usual way:

$$\begin{aligned} r_c^t &= \frac{\sum_{\mathbf{m}:\bar{\mathbf{m}}_c} \prod_{s \in \mathcal{S}} \left(\prod_{\tau=\tau_n+1}^t (1 - P_s^\tau) \cdot \tilde{P}_s^{\tau_n} \right) \cdot \prod_{s \in \bar{\mathcal{S}}} (1 - \tilde{P}_s^{\tau_n})}{\sum_{\mathbf{m}:\bar{\mathbf{m}}_c} \prod_{s \in \mathcal{S}} \left(\prod_{\tau=\tau_n+1}^t (1 - P_s^\tau) \cdot \tilde{P}_s^{\tau_n} \right) \cdot \prod_{s \in \bar{\mathcal{S}}} (1 - \tilde{P}_s^{\tau_n})} \\ &= \frac{\prod_{\tau=\tau_n+1}^t (1 - P_c^\tau) \cdot \tilde{P}_c^{\tau_n} \prod_{s \neq c} \left(\prod_{\tau=\tau_n+1}^t \cdot \tilde{P}_s^{\tau_n} + (1 - \tilde{P}_s^{\tau_n}) \right)}{\left(1 - \tilde{P}_c^{\tau_n} \right) \prod_{s \neq c} \left(\prod_{\tau=\tau_n+1}^t \cdot \tilde{P}_s^{\tau_n} + (1 - \tilde{P}_s^{\tau_n}) \right)} \end{aligned} \quad (\text{D.3})$$

Upon cancellation of the products over $s \neq c$,

$$r_c^t = \prod_{\tau=\tau_n+1}^t (1 - P_c^\tau) \cdot r_c^{\tau_n} , \quad (\text{D.4})$$

which is valid for $\tau_n < t < \tau_{n+1}$. Furthermore, with this same assumption, independence of the posteriors continues to hold for $t < \tau_{n+1}$ by trivial modification of Proof B.3. (Replace $p(\mathbf{m})$ with $P[\mathbf{m} \mid D^{\tau_n}, \bar{D}^{\tau_n}]$.) Eq. (3.19b) then follows directly by trivial modification of the derivation of (2.21) presented in App. B.5. (Replace $p(\mathbf{m} \mid \bar{D}^t)$ with $P[\mathbf{m} \mid D^{\tau_{n+1}-1}, \bar{D}^{\tau_{n+1}-1}]$.)

D.2 Extended IP

The first case of (3.25a) follows from the substitution of D^t with $D^{\tau_F^t}$ in the derivation for (3.19a). The second case of (3.25a) is trivial because by definition the conditioning on $\tilde{\rho}_c^t$ does not change for a detection. Furthermore, assumption (3.26) evaluated at $t = \tau_n$ implies independence of the posteriors conditioned on $D^{\tau_F^t}$ and \bar{D}^t continues to hold while $\mathcal{G}_F^t = \mathcal{G}_F^{\tau_n}$.

Eq. (3.25b) may then be attained by modification of the procedure used to attain the CID algorithm as follows:

$$\begin{aligned}
 r_c^t &= \frac{\text{P}[m_c \mid D^t, \bar{D}^t]}{\text{P}[\bar{m}_c \mid D^t, \bar{D}^t]} \\
 &= \frac{\text{P}[D^{\tau^t \setminus \tau_F^t} \mid D^{\tau_F^t}, \bar{D}^t, m_c]}{\text{P}[D^{\tau^t \setminus \tau_F^t} \mid D^{\tau_F^t}, \bar{D}^t, \bar{m}_c]} \cdot \frac{\text{P}[m_c \mid D^{\tau_F^t}, \bar{D}^t]}{\text{P}[\bar{m}_c \mid D^{\tau_F^t}, \bar{D}^t]} \\
 &= \frac{\text{P}[D^{\tau^t \setminus \tau_F^t} \mid D^{\tau_F^t}, \bar{D}^t, m_c]}{\text{P}[D^{\tau^t \setminus \tau_F^t} \mid D^{\tau_F^t}, \bar{D}^t, \bar{m}_c]} \cdot \tilde{\rho}_c^t.
 \end{aligned} \tag{D.5}$$

By the modified CIM assumption (3.24),

$$\begin{aligned}
 r_c^t &= \tilde{\rho}_c^t \cdot \prod_{k \in \mathcal{G}^t \setminus \mathcal{G}_F^t} \frac{\text{P}[D^{g_k^t} \mid D^{\tau_F^t}, \bar{D}^t, m_c]}{\text{P}[D^{g_k^t} \mid D^{\tau_F^t}, \bar{D}^t, \bar{m}_c]} \\
 &= \tilde{\rho}_c^t \cdot \prod_{k \in \mathcal{G}^t \setminus \mathcal{G}_F^t} \frac{\sum_{\mathbf{m}: m_c} \text{P}[D^{g_k^t} \mid \mathbf{m}] \text{P}[\mathbf{m} \mid D^{\tau_F^t}, \bar{D}^t]}{\sum_{\mathbf{m}: \bar{m}_c} \text{P}[D^{g_k^t} \mid \mathbf{m}] \text{P}[\mathbf{m} \mid D^{\tau_F^t}, \bar{D}^t]}.
 \end{aligned} \tag{D.6}$$

The latter ratio in the last line of (D.6) factors into products over $s \in \mathcal{S}$ and $s \in \bar{\mathcal{S}}$ so that the same procedure used to derive (2.21) can be employed to arrive at (3.25b).

Appendix E

Connections to Pang, 2004

This appendix explores the connections between the occupancy grid (OG) mapping methods developed in Part I of this thesis and previous work carried out by Pang [101,102] on chemical plume source localization where the search domain was similarly discretized. The essential difference is that the Bayesian map update algorithm proposed by Pang requires assuming a single source domain. This is a powerful constraint that allows inferences to be drawn about portions of the map not actually ever observed, a property that is fundamentally incompatible with multiple-source domains where the number sources cannot be known a priori.

The plume detection model developed by Pang can be extended to conform to the special form of (2.11), making it suitable for multiple source domains. This appendix presents simulation results that compare Pang’s algorithm with the recursive form of the IP algorithm developed in § 3.2.1. In a known single-source domain, Pang’s algorithm generally produces superior results. However, in multiple source domains Pang’s algorithm produces results inconsistent with the true source locations. In either case, the recursive IP algorithm produces acceptable results with approximately equivalent computational cost.

Key Probabilistic Quantities

I begin by recapitulating Pang’s model [101] for plume evolution and expressing the key probabilistic quantities in my notation. The model is appropriate to short time-scale plumes composed of approximately independently diffusing “filaments” of plume

effluent [8,30]. The key probabilistic quantity attained from this perspective on plume evolution is, in the notation of [101], $S_{ij}(t_l, t_k)$ which denotes the probability that a source in cell i released a single chemical filament at time t_l given that the filament was present in cell j at time t_k . Considered over a range of release times $t_l \in [t_0, t_k)$, the $S_{ij}(t_l, t_k)$ enable the computation of relevant probabilities for a continuously releasing source: $\omega_{ij}(t_0, t_k)$, the probability that there is a source in cell i given that there is detectable chemical in cell j at time t_k ; $\gamma_{ij}(t_0, t_k)$, the probability of not detecting a chemical in cell j at time t_k due to the continuous release of chemical from a source in cell i . For filaments released at discrete times $t_i, i \in [0, k)$, these are [101]:

$$\begin{aligned}\omega_{ij}(t_0, t_k) &= \frac{1}{k} \sum_{l=0}^{k-1} S_{ij}(t_l, t_k) \\ \gamma_{ij}(t_0, t_k) &= \prod_{l=0}^{k-1} (1 - \mu S_{ij}(t_l, t_k))\end{aligned}\tag{E.1}$$

where μ denotes a detection probability given that both the sensor and a chemical filament occupy the same cell. Note that for cells not upwind of cell j , $\omega_{ij} \rightarrow 0$ and $\gamma_{ij} \rightarrow 1$, which again reflects the assumption of a single source. In my notation, these are

$$\begin{aligned}\text{P} [\mathcal{S} = i \mid d^k, S = 1] &= \omega_{ij}(t_0, t_k) \\ \text{P} [\bar{d}^k \mid \mathcal{S} = i] &= \gamma_{ij}(t_0, t_k)\end{aligned}\tag{E.2}$$

where the sensor implicitly occupies cell j at time t_k . These probabilities represent respectively half of an inverse model, and half of a forward model.

Update Algorithm

Pang describes a recursive update rule using these quantities for the probability that the source lies in cell i . Because a single source is assumed, the quantity being estimated for each cell is $\text{P} [\mathcal{S} = i \mid \Delta^t, S = 1]$ and not $\text{P} [m_i \mid \Delta^t]$ as in occupancy grid mapping. In my notation, his update rules for a detection and non-detection

registered at time t_k are, respectively:

$$\begin{aligned} P[\mathcal{S} = i \mid d^k, \Delta^{k-1}] &= \frac{P[\mathcal{S} = i \mid d^k]}{P[\mathcal{S} = i]} P[\mathcal{S} = i \mid \Delta^{k-1}] \\ P[\mathcal{S} = i \mid \bar{d}^k, \Delta^{k-1}] &= \frac{P[\bar{d}^k \mid \mathcal{S} = i]}{P[\bar{d}^k]} P[\mathcal{S} = i \mid \Delta^{k-1}] , \end{aligned} \tag{E.3}$$

where the conditioning of all quantities on $S = 1$ remains but has been dropped for brevity.

An independence assumption is required to derive (E.3): $p(\delta^k, \Delta^{k-1}) = P[\delta^k] P[\Delta^{k-1}]$. The update rules that result have intuitive appeal and simulation results in [101] indicate their utility for the problem studied. If $P[\mathcal{S} = i \mid d^k] > P[\mathcal{S} = i]$, then one expects a detection given a source in i , and the posterior probability that cell i contains the source increases. Additionally, because $P[\mathcal{S} = i \mid d^k]$ approaches 0 for cells not upwind of i rather than approaching the prior, the posteriors of cells not upwind are reduced. Likewise, if $P[\bar{d}^k \mid \mathcal{S} = i] < P[\bar{d}^k]$, then one expects a detection given a source in cell i , and since none was registered, the posterior for cell i is decreased. As above, $P[\bar{d}^k \mid \mathcal{S} = i]$ does not approach the prior for cells not upwind, instead approaching 1, and consequently those cells will have their posteriors increased. This behavior is entirely consistent with the assumption of a single source in the search domain; however, it allows inferences to be made about the entire map regardless of the portion of the map actually observed, and this property is fundamentally incompatible with multiple-source domains.

The derivation of (E.3) also requires that $p(\delta^k, \Delta^{k-1} \mid \mathcal{S} = i) = p(\delta^k \mid \mathcal{S} = i)$. This relationship is approximately true for all upwind and cross-wind sensor trajectories assuming the source location is static. It is, in fact, the static world assumption (§ 2.1.1) stated for all single-source maps and should not be confused with the CIM assumption required by standard OG methods.

Adaptation for Multiple-Source Domains

Pang's model can be adapted to multiple source domains by recognizing that the $S_{ij}(t_l, t_k)$ are related to the probability of non-detection given a single source map by

the second part of (E.1).¹ Restating (2.12),

$$P [\bar{d}^k \mid \mathcal{S} = i] = (1 - P_F^k)(1 - P_i^k) \quad (\text{E.4})$$

then leads immediately to

$$P_i^k = 1 - \frac{\gamma_{ij}(t_0, t_k)}{1 - P_F^k} . \quad (\text{E.5})$$

The P_i^k are the fundamental probabilities required to implement any of the OG mapping algorithms developed in this work.

Simulation Results

Figure E-1 shows the results of Pang’s algorithm, the recursive IP algorithm, and the exact algorithm (§ 2.4) applied to a small single-source domain simulation. Figure E-2 shows these same algorithms applied to a domain with two sources. The simulated plume (black) and vehicle trajectory (blue, red) is shown at four times during the simulation in the bottommost panels of each figure. Non-detections are indicated by blue dots; detections by red dots. The remaining panels show the outputs of each algorithm. Cell colors in the results for Pang’s algorithm approximate the \log_{10} odds that the correct map has the corresponding cell occupied and all others empty. Cell colors in the recursive IP results approximate the \log_{10} odds of the marginal posterior of that cell being occupied. The correct marginal posterior odds are shown in the topmost column of panels in each figure.

Pang’s algorithm produces superior results when applied to a single-source domain but not when applied to a domain with multiple sources. Both the superior results in the former case and the algorithm’s behavior in the latter stem from the assumption of a single source. In a single-source domain, the assumption of a single source enables inferences to be drawn about portions of the map not actually observed during a specific measurement. From the figures, before the first detection, Pang’s algorithm gradually raises the odds in portions of the map not observed by the non-detections. Upon registering the first detection, the odds in all of the map not immediately upwind

¹The quantity $S_{ij}(t_l, t_k)$ equivalently specifies both a forward and inverse model. It is at once both the probability that a filament detected in cell j was released from cell i (inverse), and that a filament released from cell i is in cell j (forward).

are dropped to near zero. This causes problems when another detection suggests a source within one of these regions (Fig. E-2): all data is interpreted in terms of a single source so that widely spaced detections are interpreted as evidence of a single source far upwind. It is important to note that Pang's algorithm computes a set of probabilities that explicitly assume a single source: $P[\mathcal{S} = i \mid \Delta^k, S = 1]$. The results produced by Pang's algorithm in the multiple-source domain simulation are a reasonable result for the data if the environment were in fact known to have one source. However, many practical scenarios do not allow such an assumption to be made. In both figures, the recursive IP results are similar to the correct marginal posterior odds computed by the exact algorithm; though the portions of the map observed during early detection times but not observed later has somewhat high odds relative to the exact result, a consequence of the uni-directional linkage between past and current measurements that is characteristic of the IP class of algorithms.

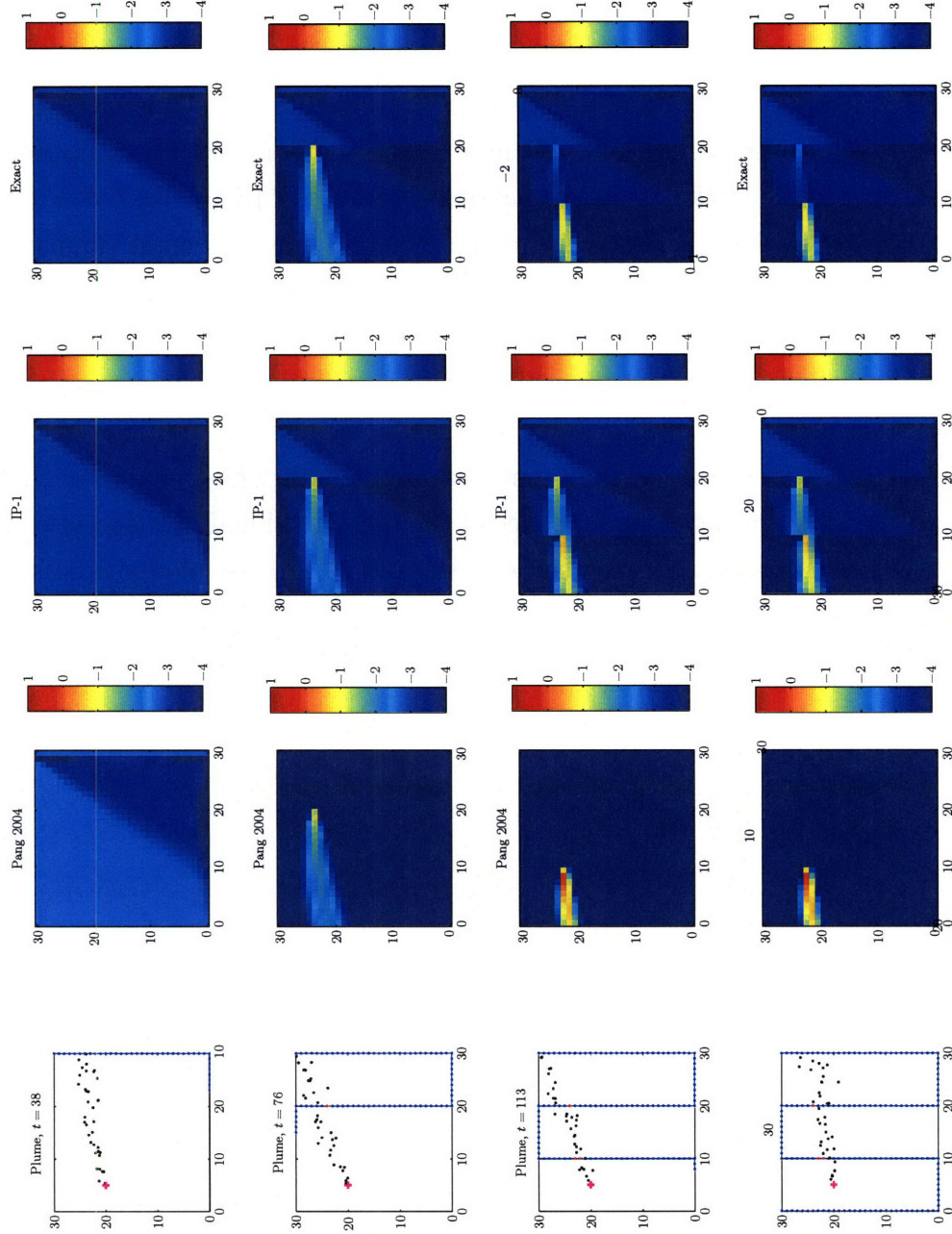


Figure E-1: Pang’s [101, 102] algorithm vs. the recursive IP and exact OG mapping algorithms in a single-source domain. See the main text for an explanation of each panel. The results for Pang’s algorithm are superior in the sense that most of the map has been correctly identified empty and with only a single peak surrounding the correct source location. In contrast, the other algorithms have produced results with higher posterior odds over much of the map, though the largest peak in each case occurs in the right location. The difference is a consequence of the former algorithm assuming a known number of sources which enables it to infer information about all portions of the map, not just those directly upwind.

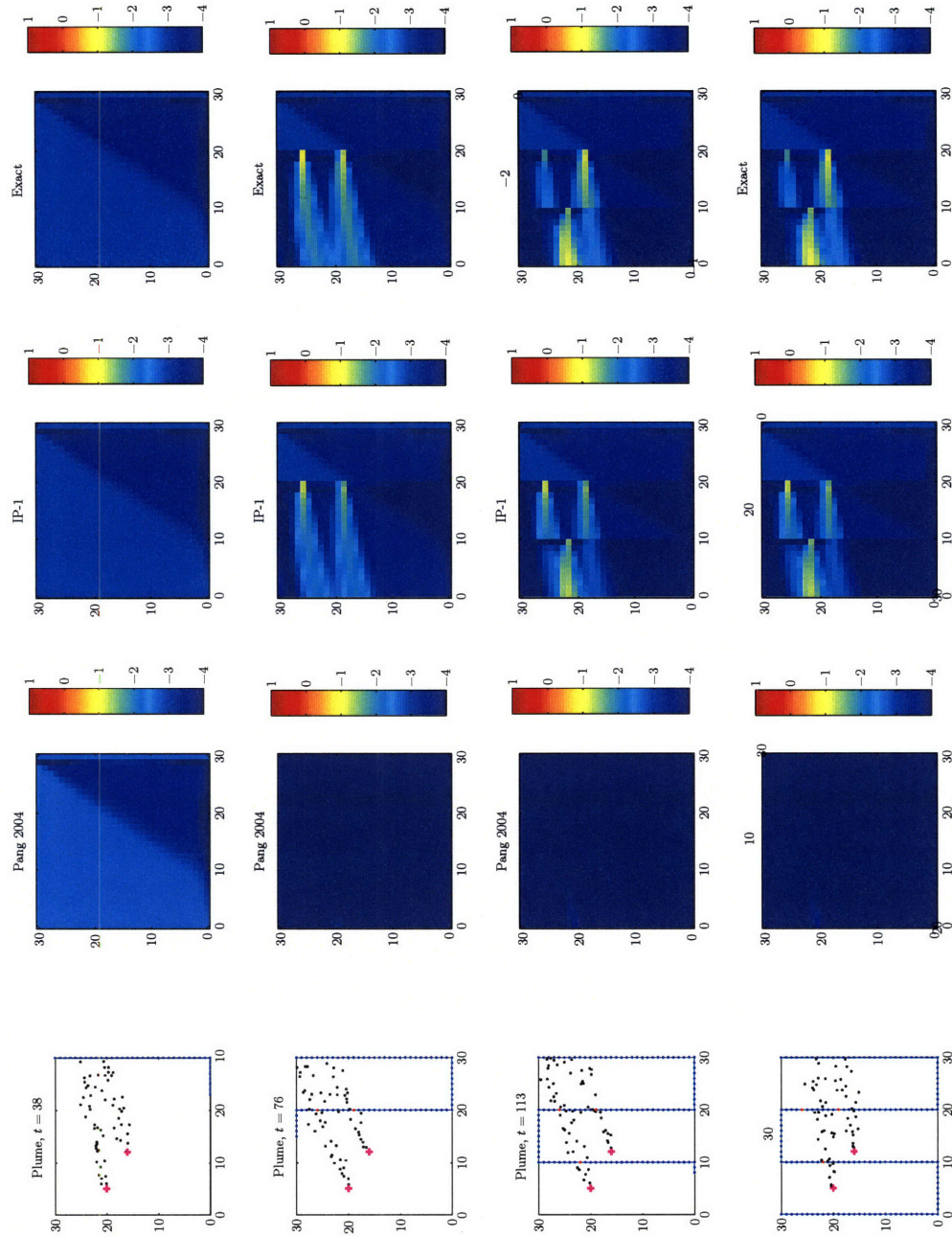


Figure E-2: Pang's [101,102] algorithm vs. the recursive IP and exact OG mapping algorithms in a multiple-source domain. See the main text for an explanation of each panel. The second detection is widely spaced from the first and inconsistent with a single source in the domain. Pang's algorithm produces a small peak away from any of the the true sources. These results improve slightly upon further detections, all from the same source; however, the single peak in the map still does not coincide with either source though it is close to the leftmost source. Both the recursive IP and exact algorithms produce multi-modal maps with relatively high posteriors odds over the true source locations. One additional high posterior odds area is also present and not associated with a true source; however, it is still consistent with the data as revealed by the exact results.

Appendix F

Anomaly Maps

This appendix is an archive of hydrothermal vent prospecting data from the Autonomous Benthic Explorer (ABE) dives studied in this thesis. Table F.1 lists the dives studied by site. Sites at ELSC are identified by their at-sea site number designation as well as by site name.

The following sections are organized by site. Each starts with a summary plot of vent prospecting results at that site using the methods developed in Ch. 4. This is followed by plots of summary data from each Phase-1 and Phase-2 dive at the site. The plots show profiles collected during vehicle descent, time series data for each tracer (after pre-processing, processed for NBP detection, processed for BP detection), bird's eye views of anomaly intensity (processed for NBP detection, processed for BP detection), and a bird's eye view of each horizontal crossflow record from ABE's ADCP.¹ Detailed descriptions of each type of plot can be found in Ch. 4.

The most significant room for improvement lies with detection of the non-buoyant plume (NBP). As a whole the NBP time series plots in this appendix show NBP contact over more of each dive than the automatically identified intervals. Likely culprits are non-linearity of the background profile, changes in sensor behavior during descents versus horizontal surveying, and horizontal variability in the background profile. In many cases the true background appears somewhat offset from the center of the prediction interval indicating that simply reducing the probability of false alarm would not yield improved performance. Detection of NBP would benefit from batch

¹The crossflow record for ABE-134 suffers from compass calibration problems and is not included.

Table F.1: ABE dives studied in this thesis. Bottom depths are approximate averages over the course of entire dives.

Site	Dive	Phase	Bottom Depth	Depth	Height	Comments
ELSC-1	126	1	2650 m	2315 m	N/A	
ELSC-1	137	2	2650 m	N/A	50 m	
ELSC-3	128	1	2165 m	1880 m; 1960 m	N/A	two passes at different depths
ELSC-3	136	2	2165 m	N/A	50 m	
ELSC-5	131	1	2050 m	1880 m; 1800 m; 1615 m	N/A	three distinct depths over partially overlapping areas
ELSC-5	133	2	2050 m	N/A	50 m	
ELSC-5	134	2	2050 m	N/A	50 m	
SMAR	150	1	3000 m	2875 m; 2750 m	N/A	alternating depths on each trackline
SMAR	151	2	3000 m	N/A	50 m	
SMAR	153	2	3000 m	N/A	50 m	

processing of the entire record instead of the present technique which relies on the accuracy of the background profile determined during vehicle descent.

F.1 ELSC: Site-1 (Kilo Moana)

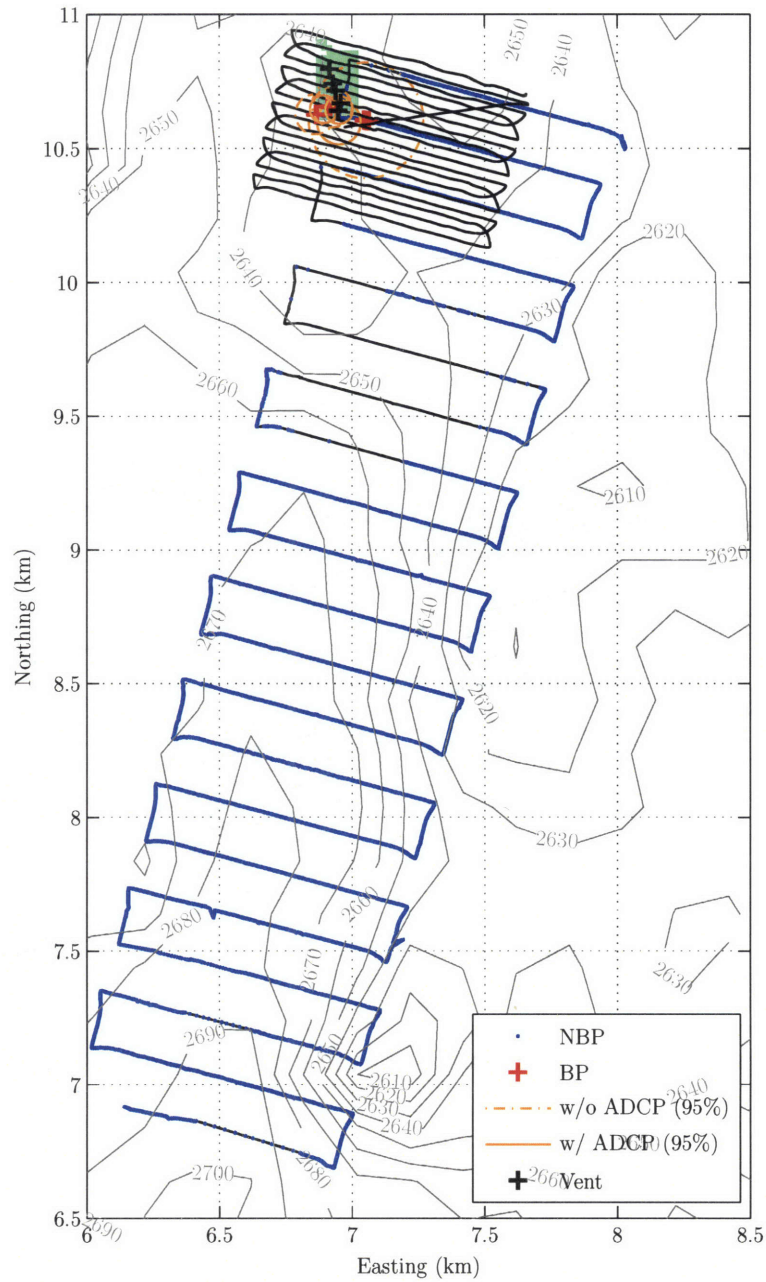


Figure F-1: Vent prospecting summary data, ELSC-1.

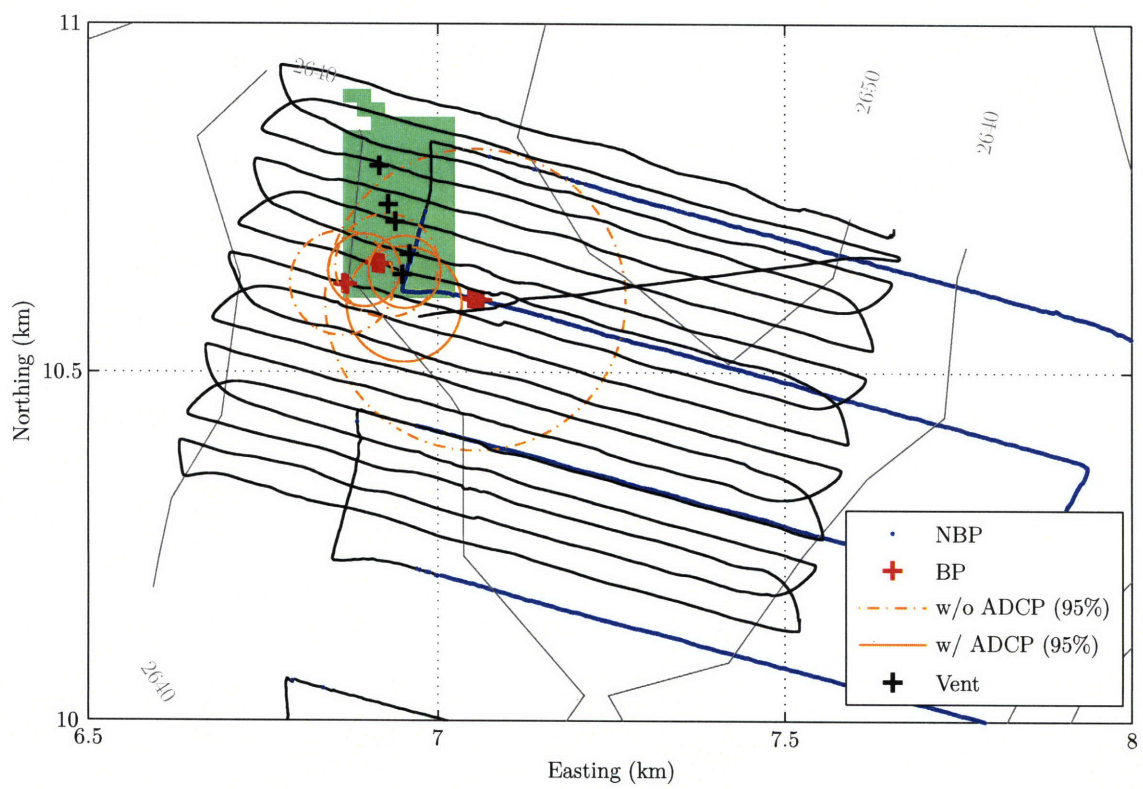


Figure F-2: Closeup of vent prospecting summary data, ELSC-1.

F.1.1 ABE-126

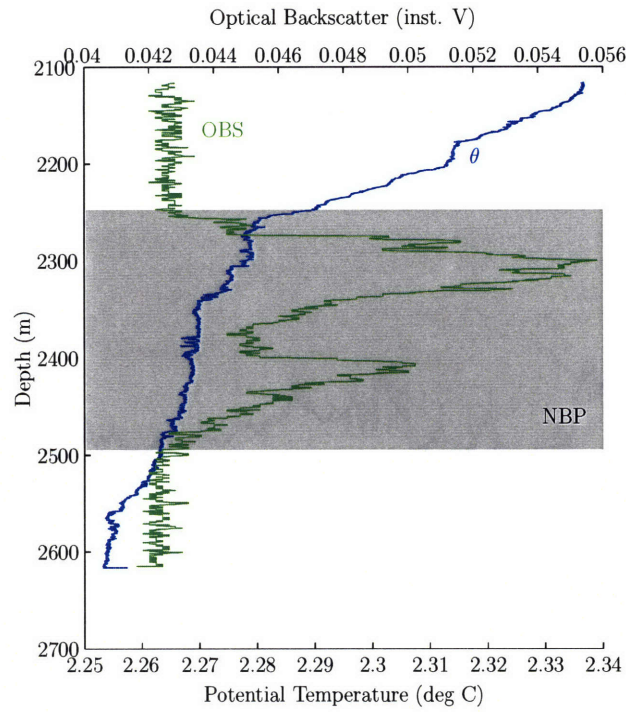


Figure F-3: Descent profiles vs. depth, ABE-126.

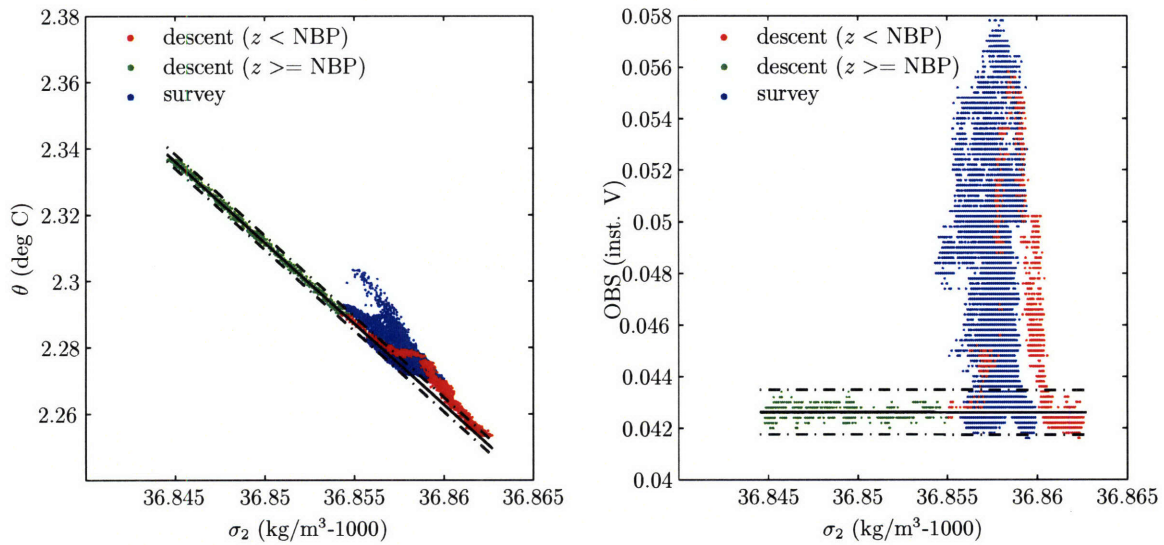


Figure F-4: Descent profiles vs. potential density, ABE-126.

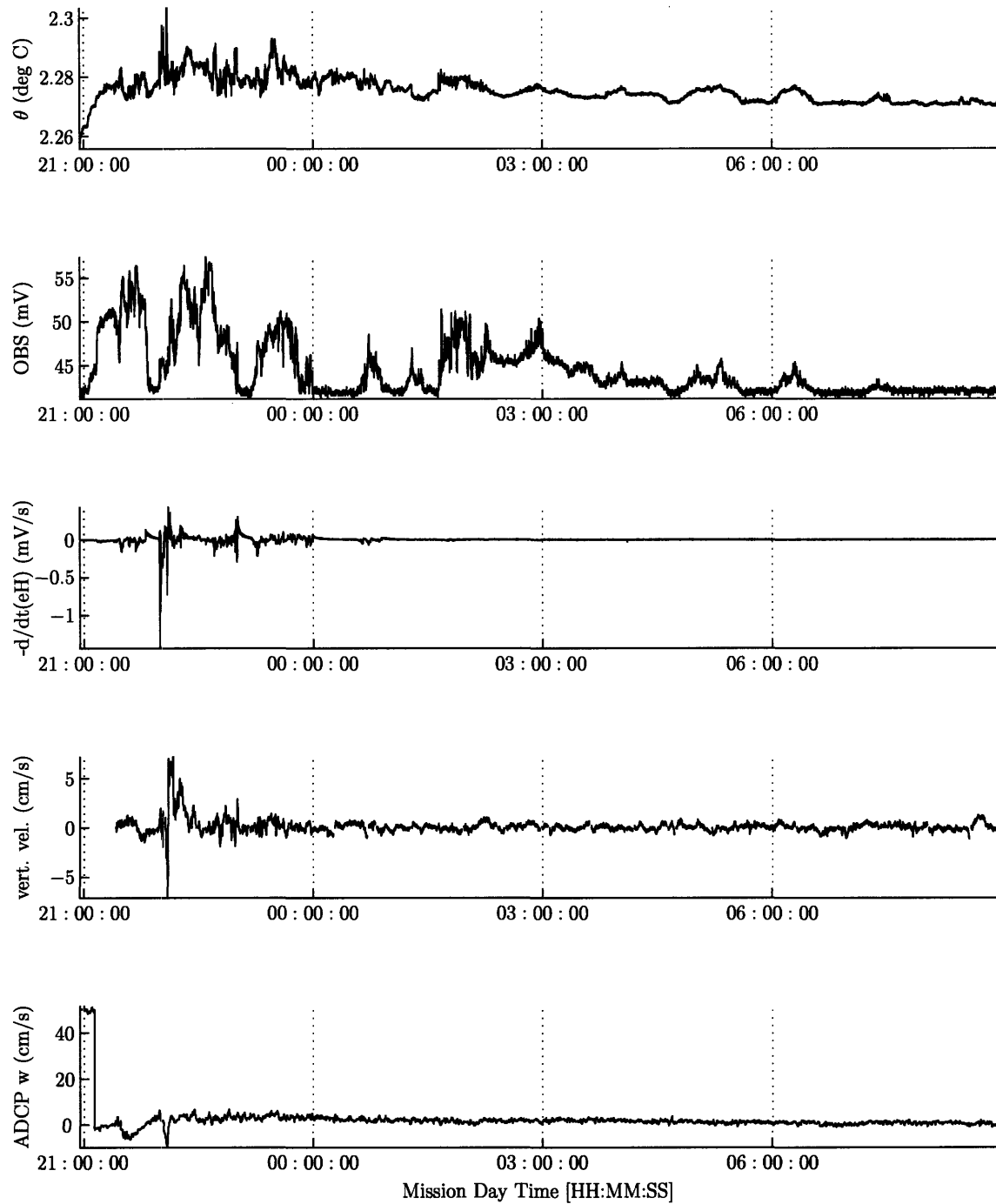


Figure F-5: Time series of hydrothermal tracers after pre-processing, ABE-126.

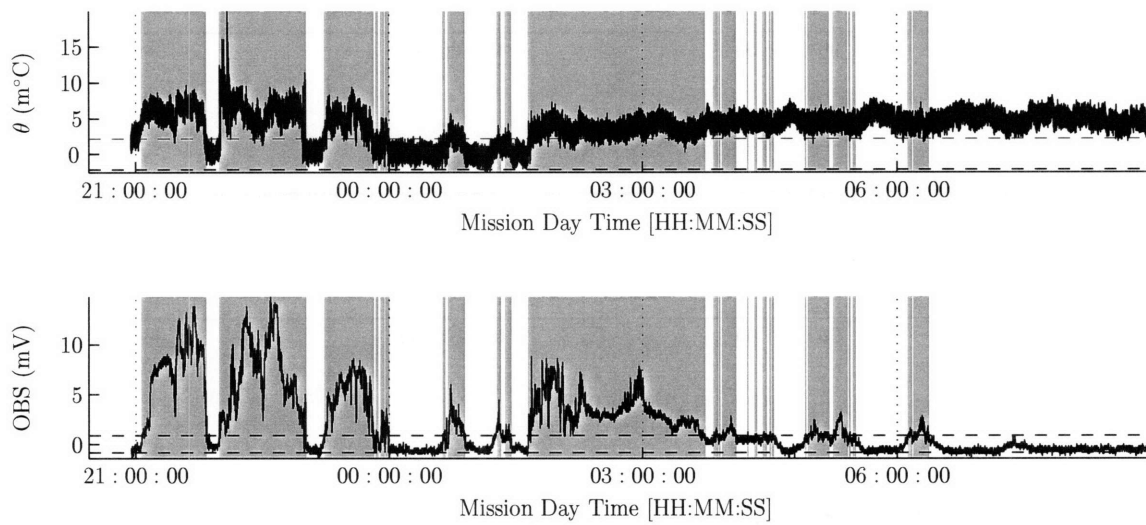


Figure F-6: Time series of hydrothermal tracers processed for NBP detection, ABE-126. Intervals highlighted in gray indicate NBP detection.

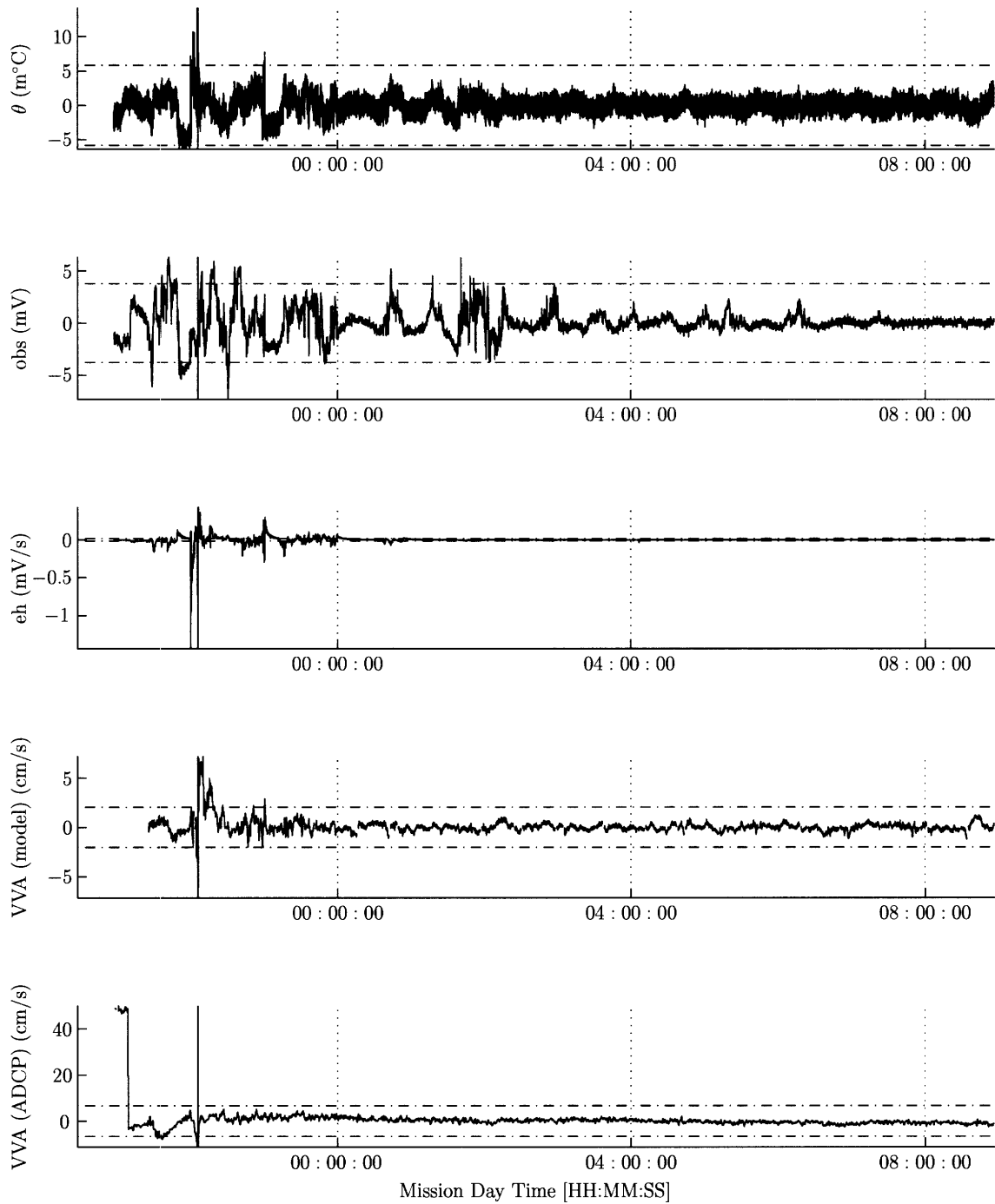


Figure F-7: Time series of hydrothermal tracers processed for BP detection, ABE-126. Intervals highlighted in gray indicate BP detection.

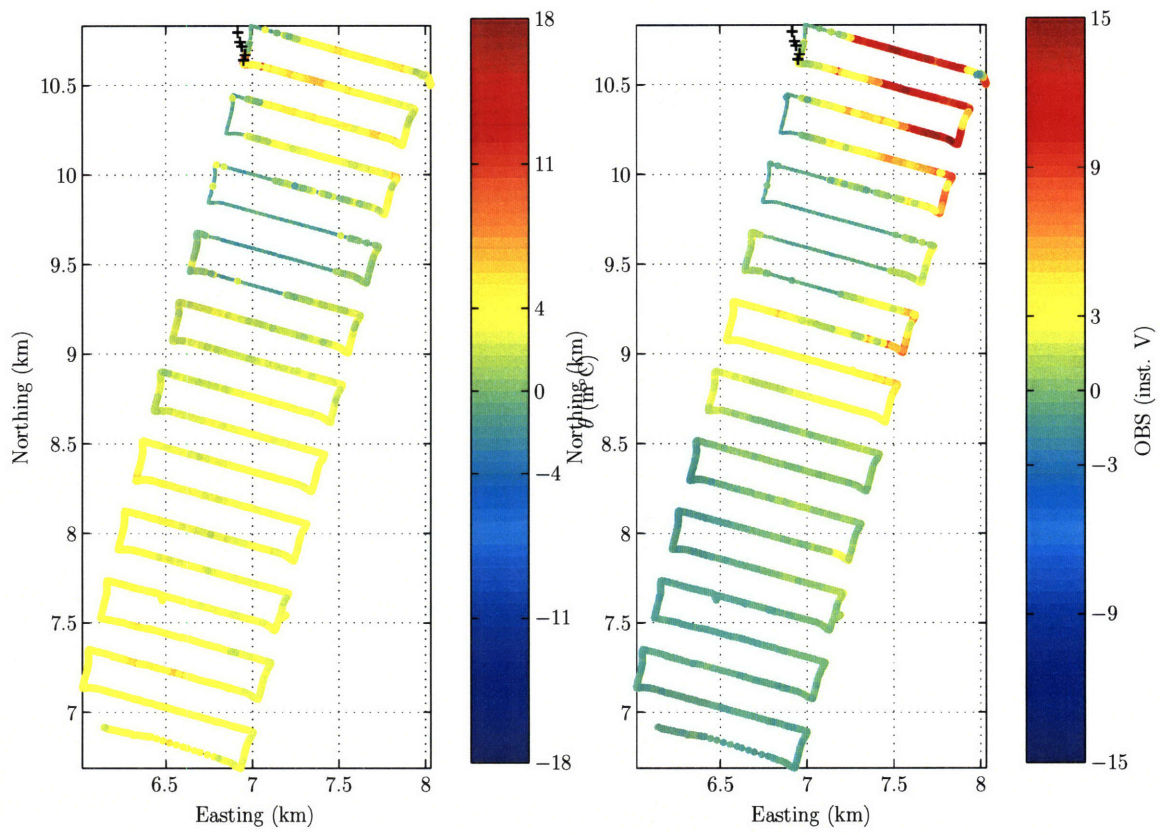


Figure F-8: Bird's eye view of anomaly intensity after processing for NBP detection, ABE-126. Larger diameter dots indicate NBP detection.

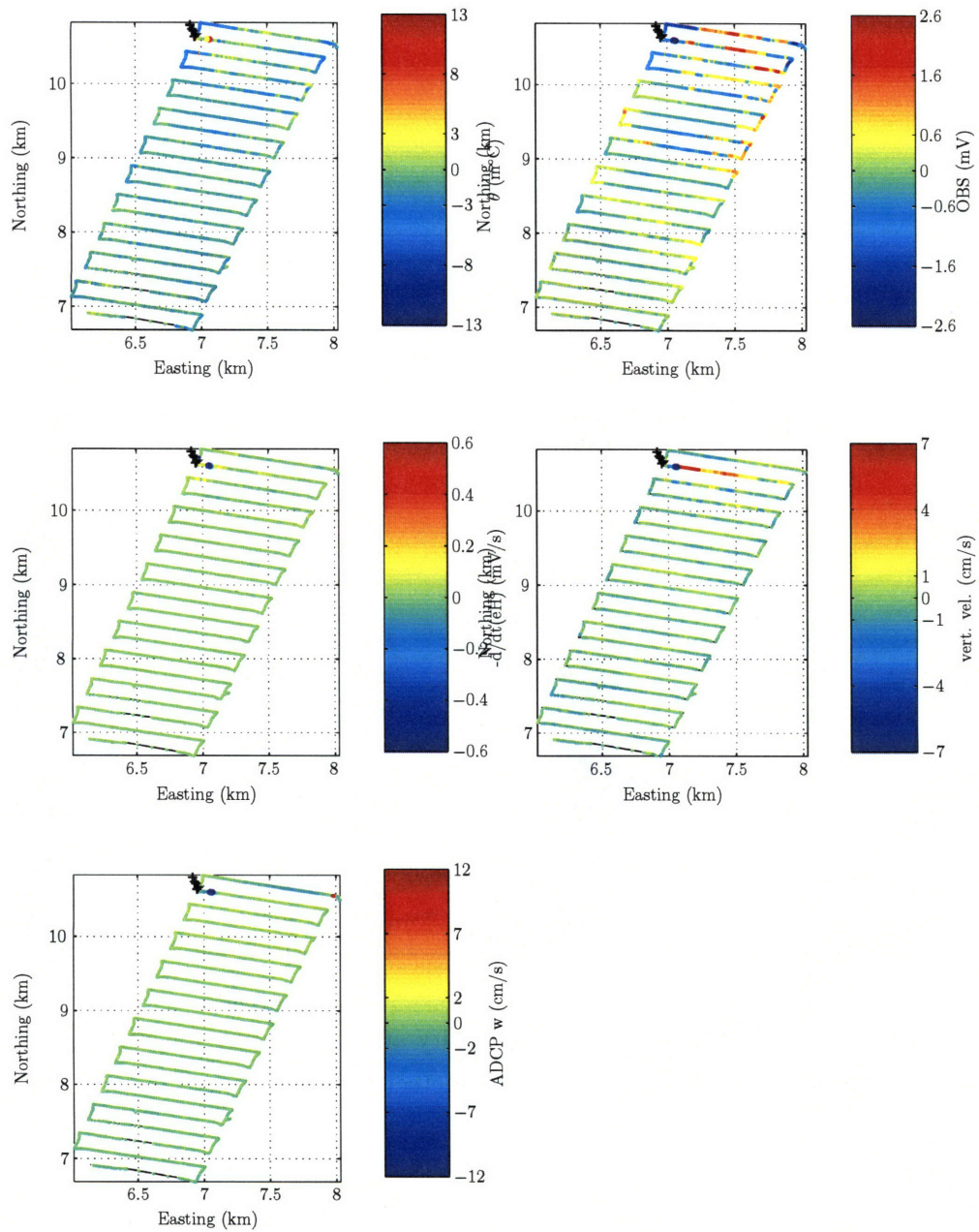


Figure F-9: Bird's eye view of anomaly intensity after processing for BP detection, ABE-126. Larger diameter dots indicate BP detection.

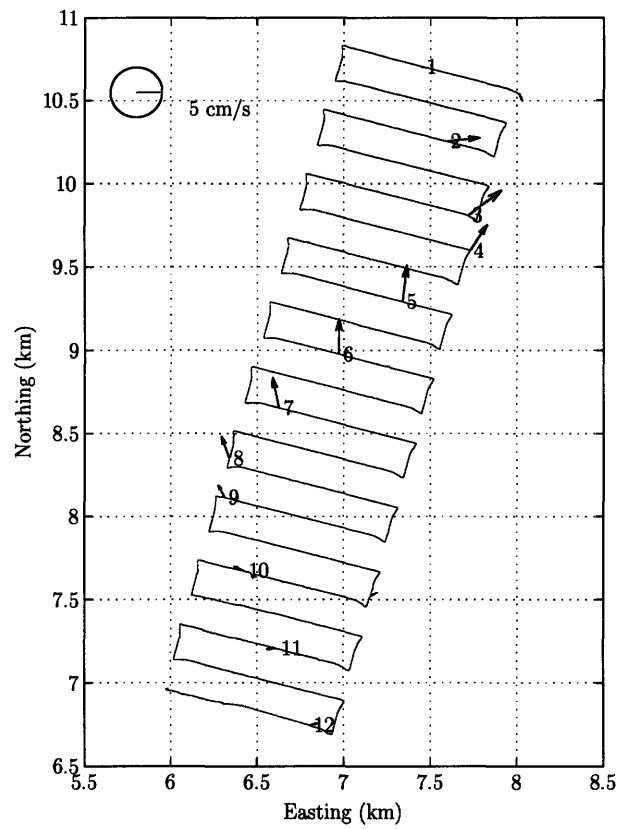


Figure F-10: Bird's eye view of ADCP-derived crossflow velocity, ABE-126 (1 h vertical average over depth interval 5 m to 25 m below vehicle). Numbers indicate the order of observation.

F.1.2 ABE-137

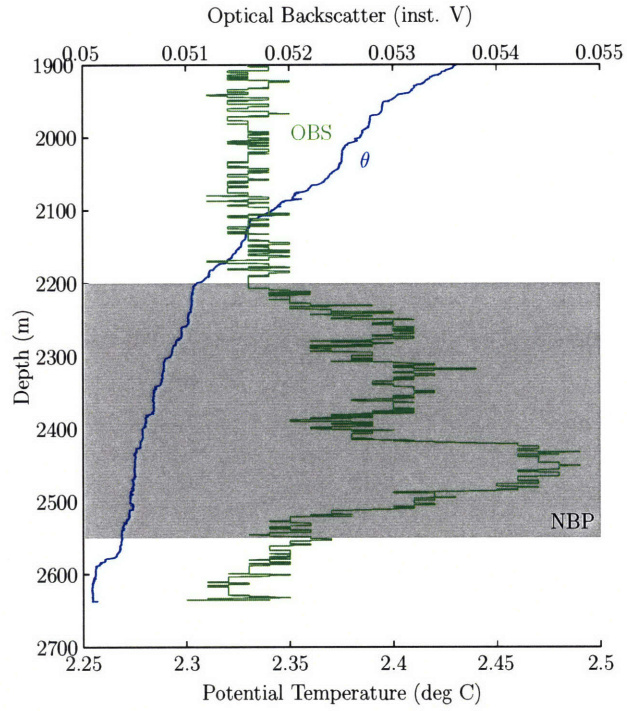


Figure F-11: Descent profiles vs. depth, ABE-137.

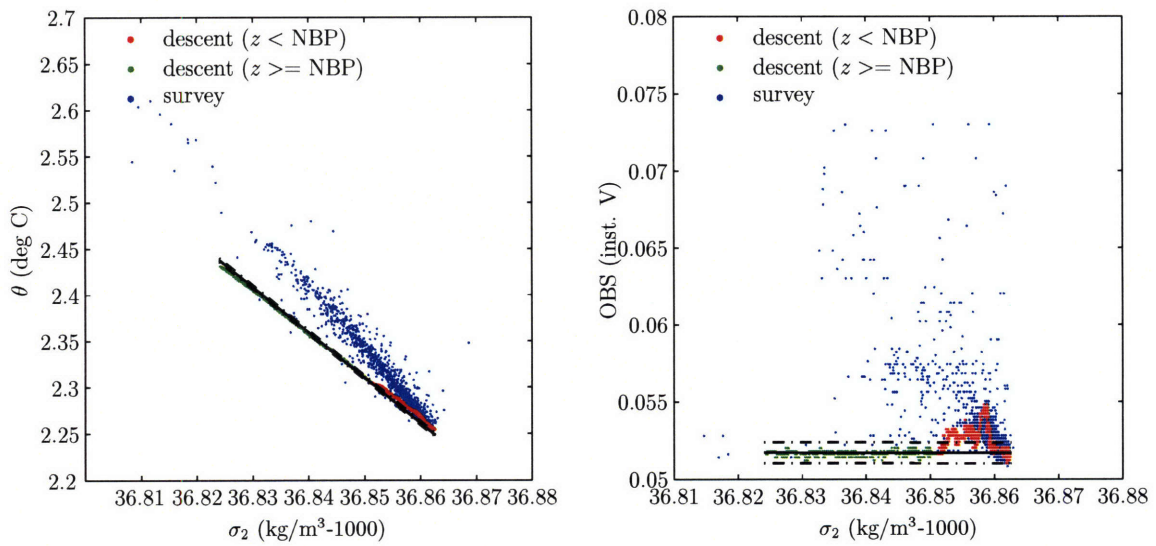


Figure F-12: Descent profiles vs. potential density, ABE-137.

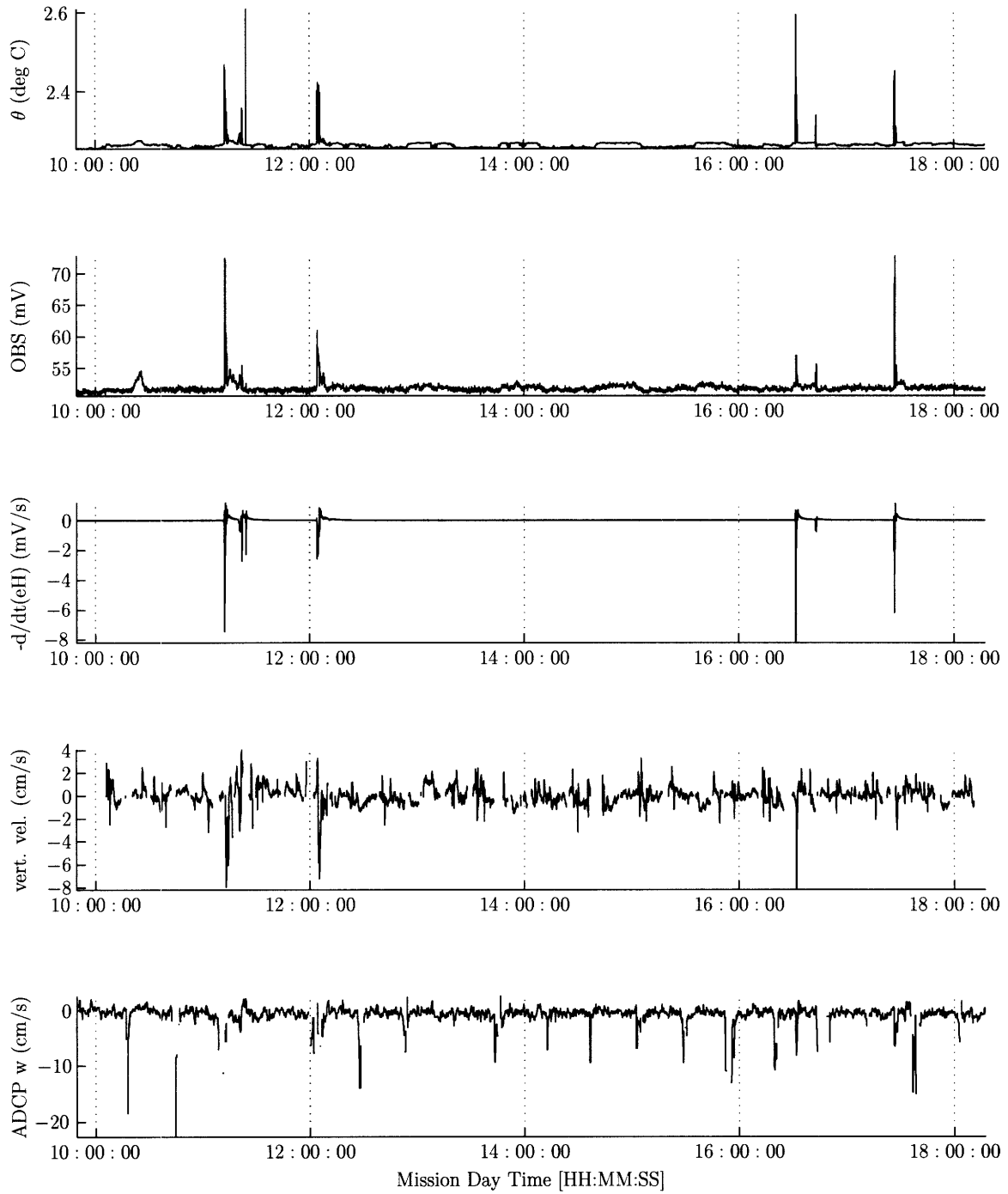


Figure F-13: Time series of hydrothermal tracers after pre-processing, ABE-137.

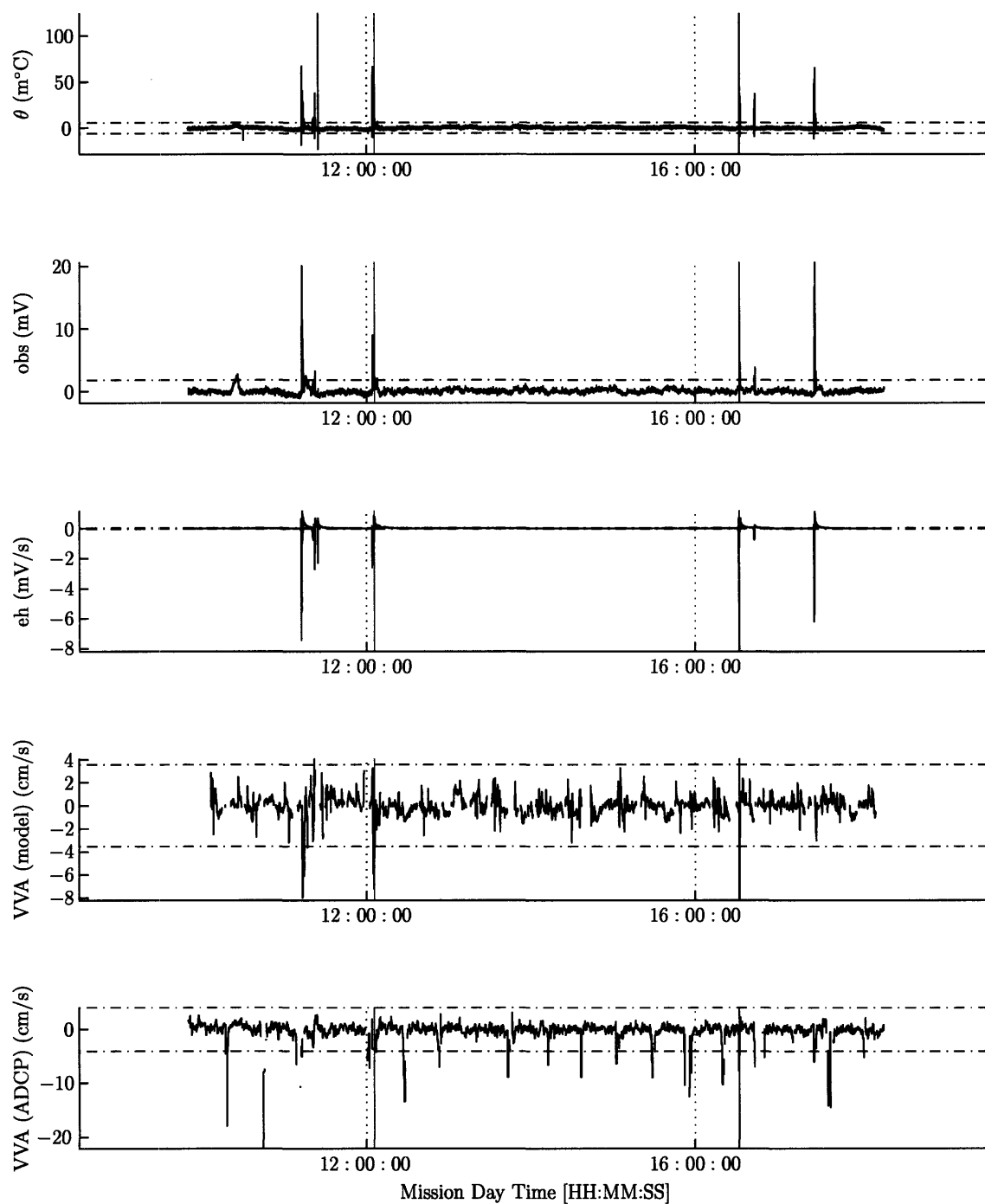


Figure F-14: Time series of hydrothermal tracers processed for BP detection, ABE-137. Intervals highlighted in gray indicate BP detection.

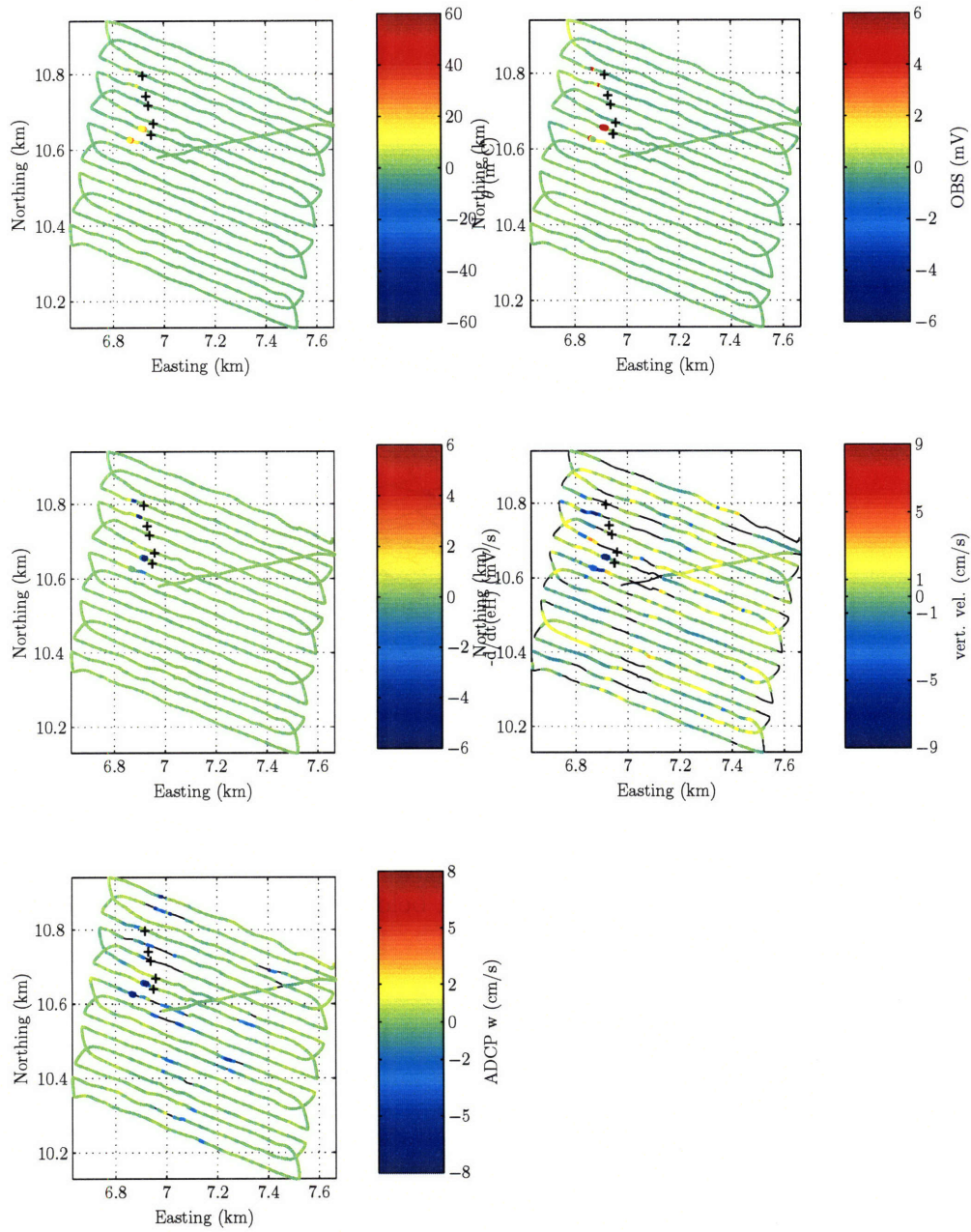


Figure F-15: Bird's eye view of anomaly intensity after processing for BP detection, ABE-137. Larger diameter dots indicate BP detection.

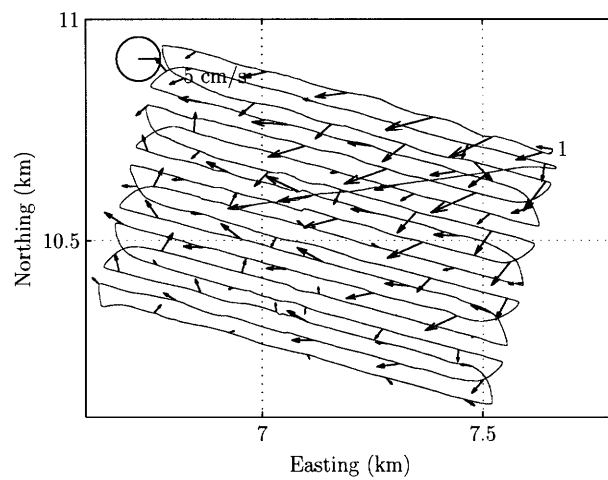


Figure F-16: Bird's eye view of ADCP-derived crossflow velocity, ABE-137 (5 min vertical average over depth interval 5 m to 25 m below vehicle). Numbers indicate the order of observation.

F.2 ELSC: Site-3 (ABE Site)

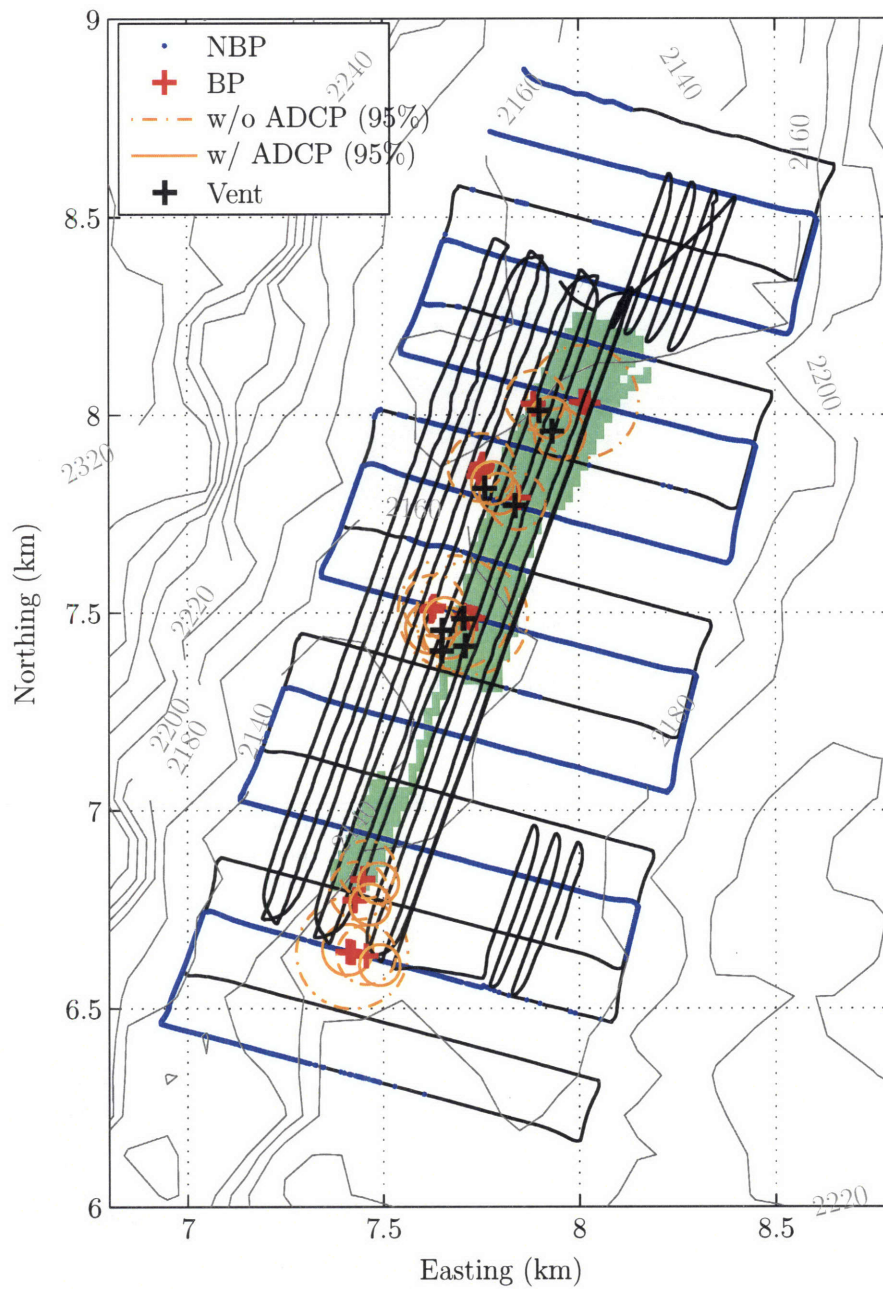


Figure F-17: Vent prospecting summary data, ELSC-3.

F.2.1 ABE-128

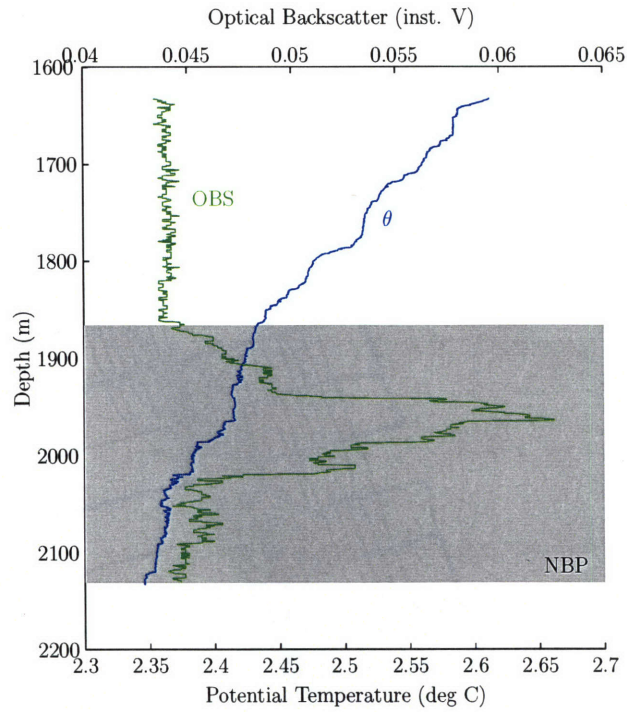


Figure F-18: Descent profiles vs. depth, ABE-128.

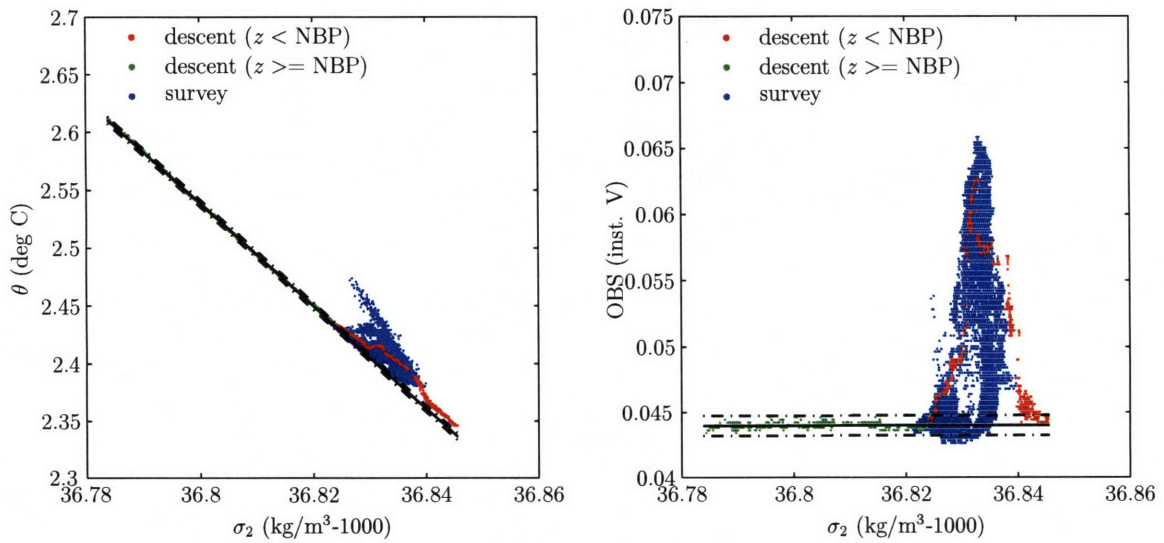


Figure F-19: Descent profiles vs. potential density, ABE-128.

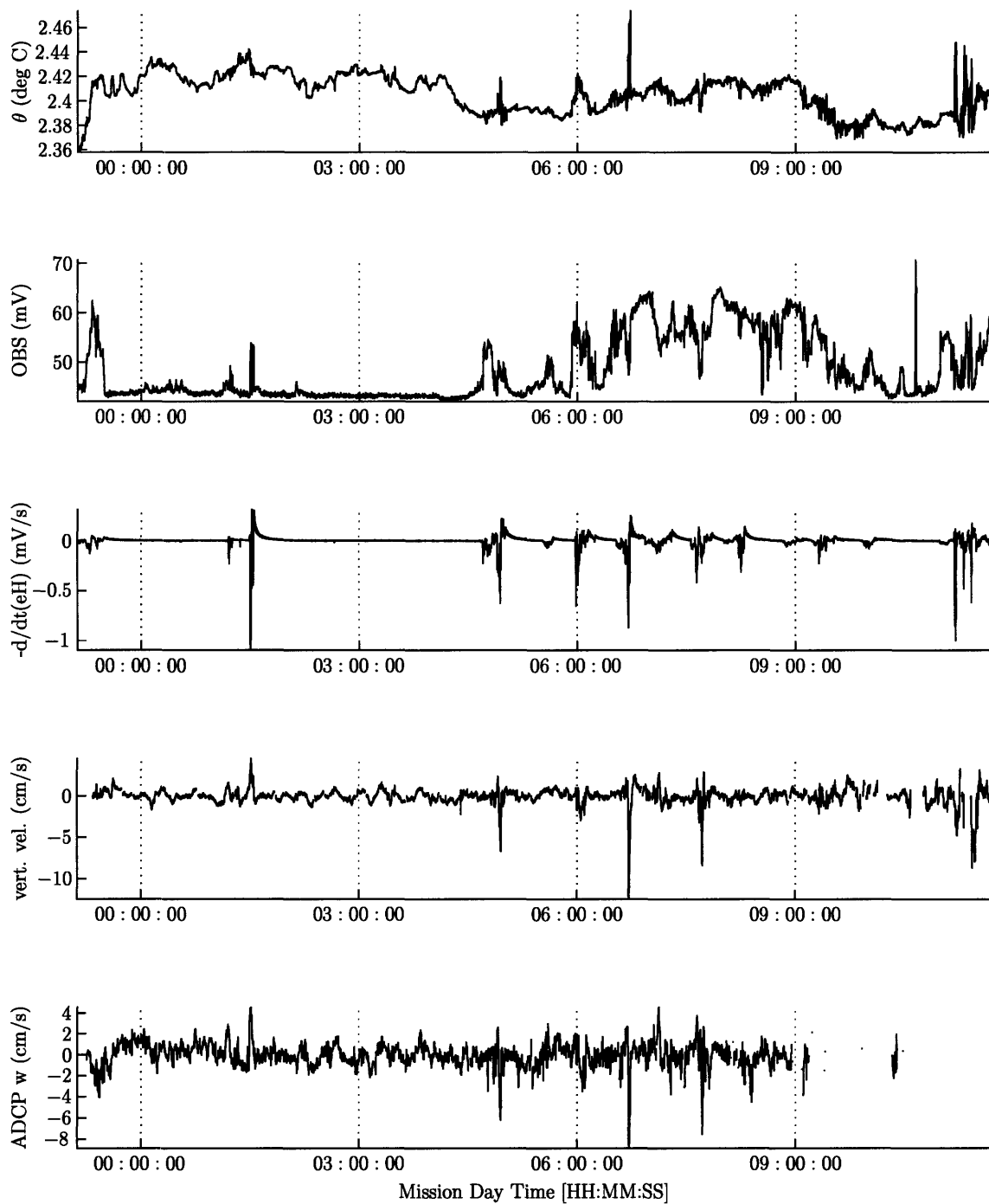


Figure F-20: Time series of hydrothermal tracers after pre-processing, ABE-128.

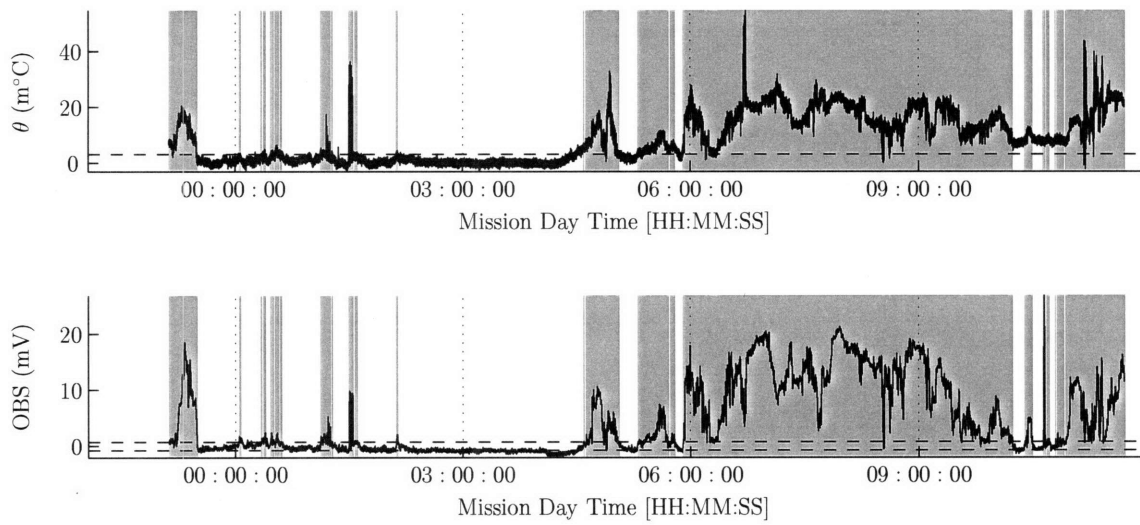


Figure F-21: Time series of hydrothermal tracers processed for NBP detection, ABE-128. Intervals highlighted in gray indicate NBP detection.

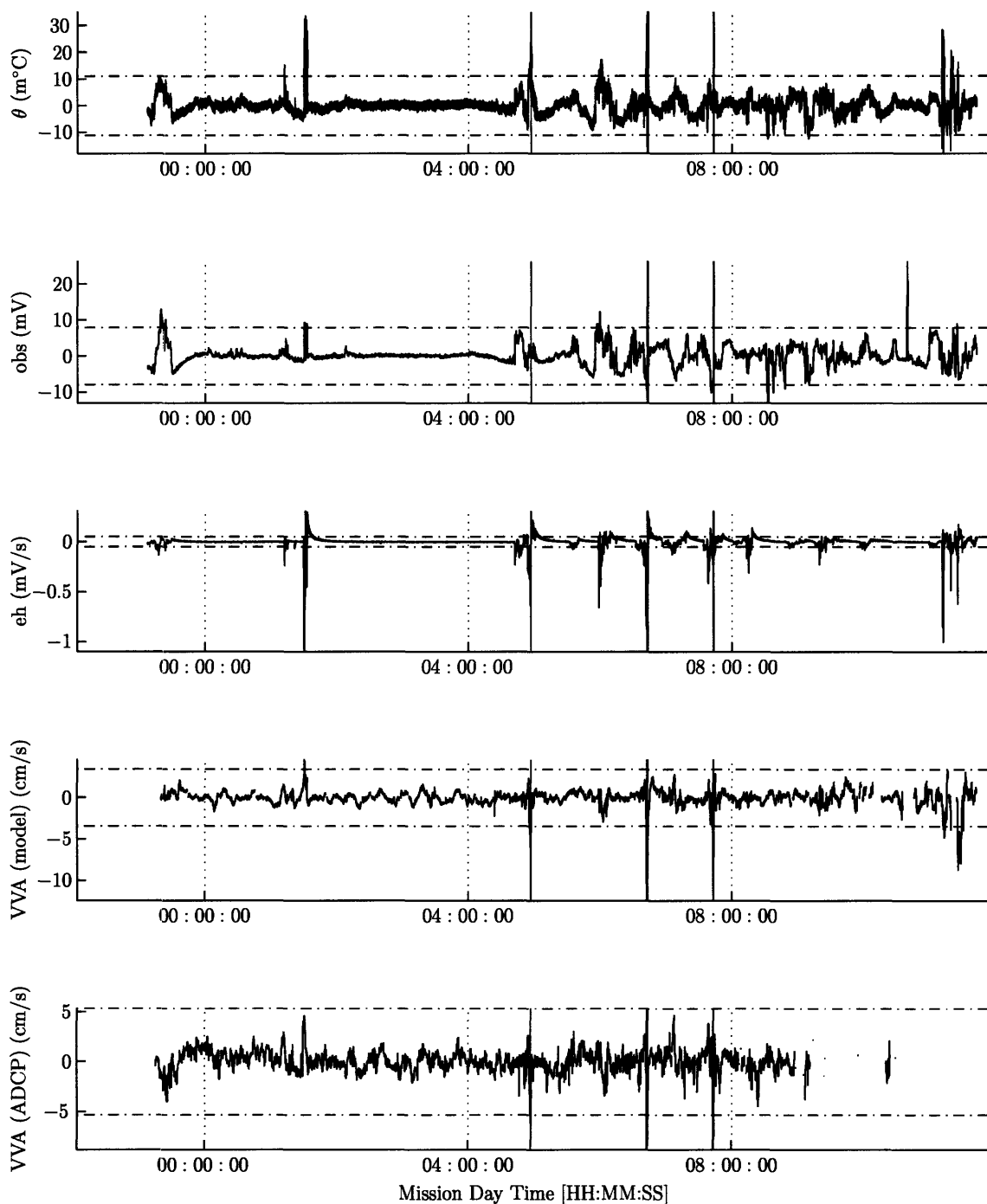


Figure F-22: Time series of hydrothermal tracers processed for BP detection, ABE-128. Intervals highlighted in gray indicate BP detection.

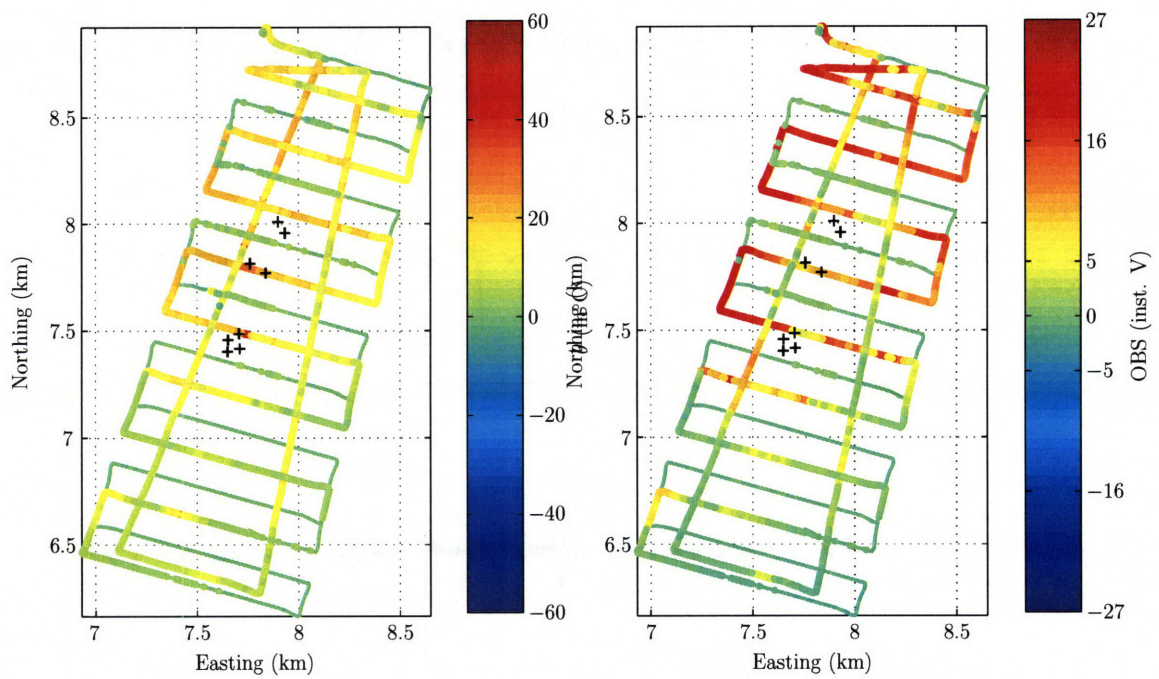


Figure F-23: Bird's eye view of anomaly intensity after processing for NBP detection, ABE-128. Larger diameter dots indicate NBP detection.

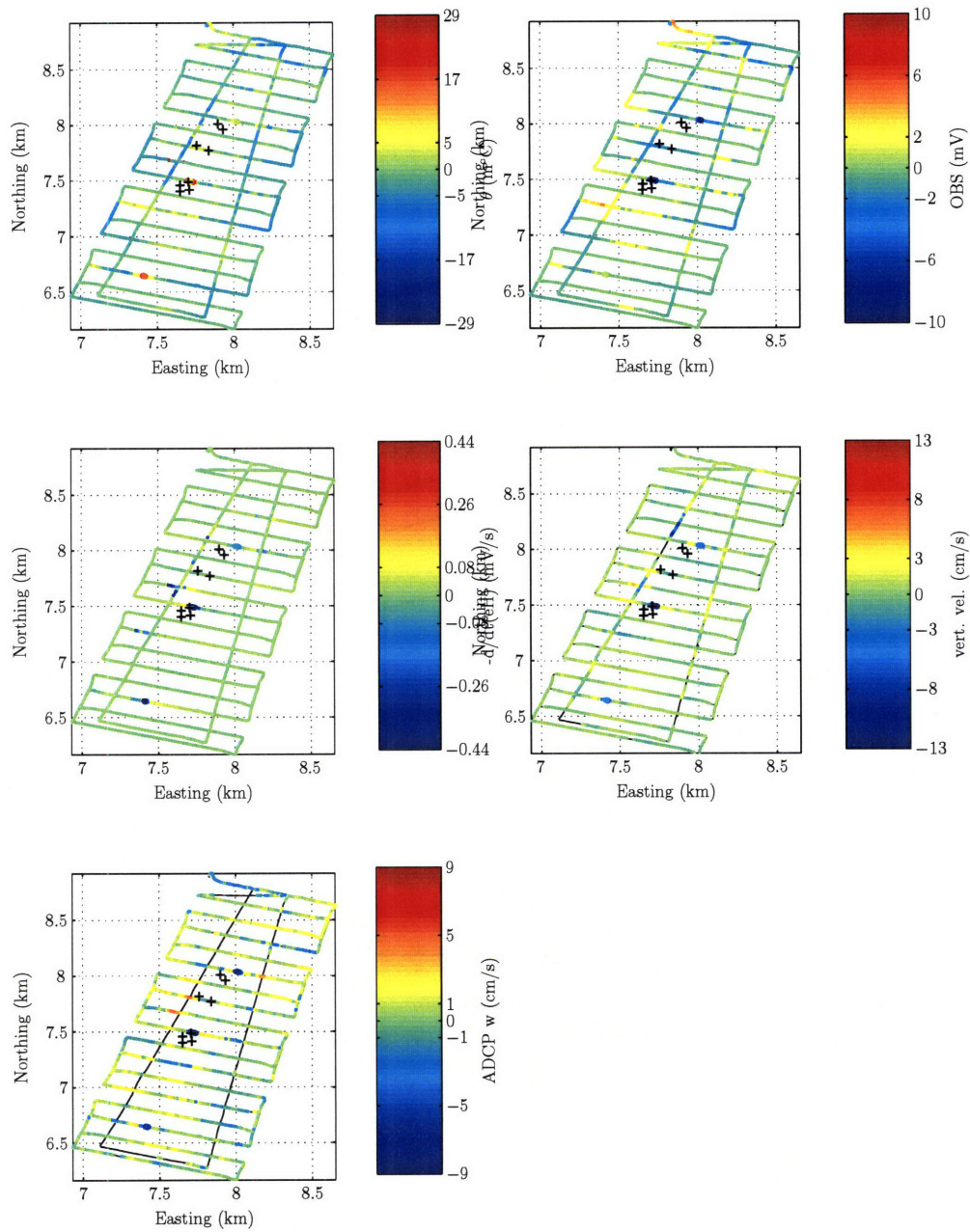


Figure F-24: Bird's eye view of anomaly intensity after processing for BP detection, ABE-128. Larger diameter dots indicate BP detection.

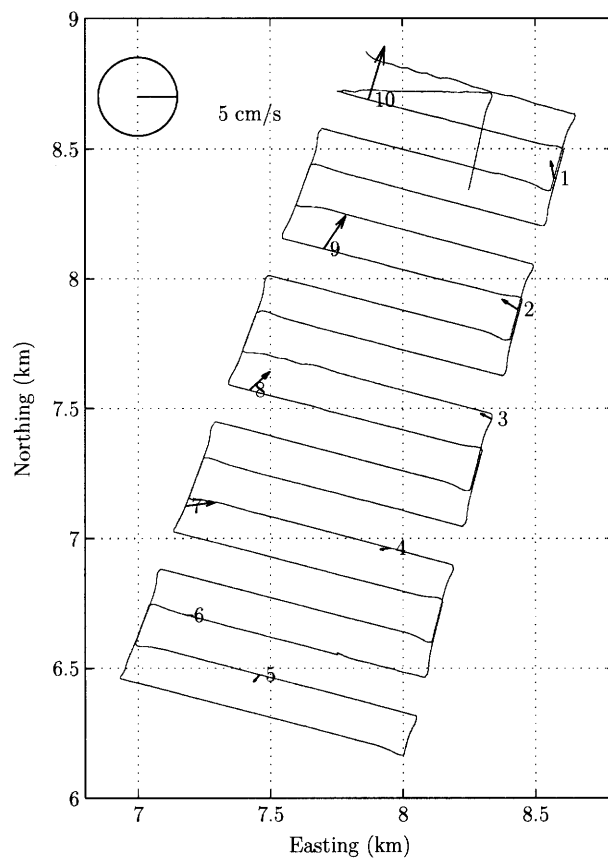


Figure F-25: Bird's eye view of ADCP-derived crossflow velocity, ABE-128 (1 h vertical average over depth interval 5 m to 25 m below vehicle). Numbers indicate the order of observation.

F.2.2 ABE-136

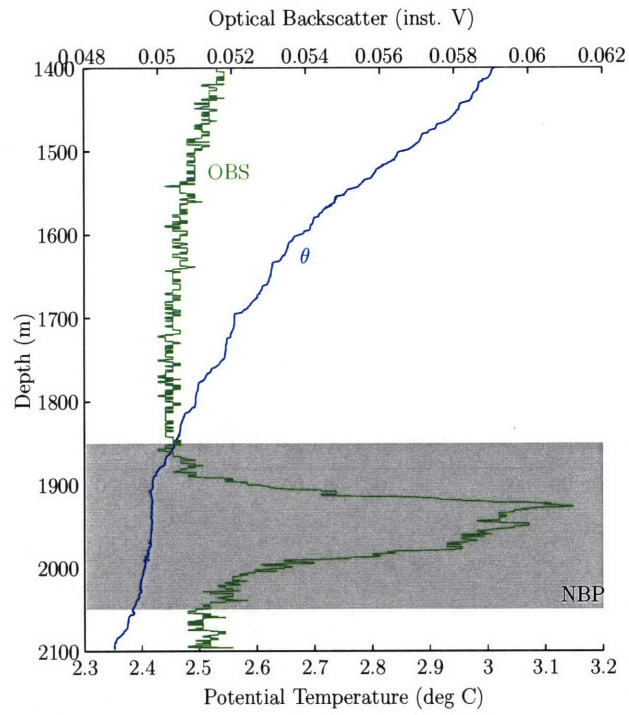


Figure F-26: Descent profiles vs. depth, ABE-136.

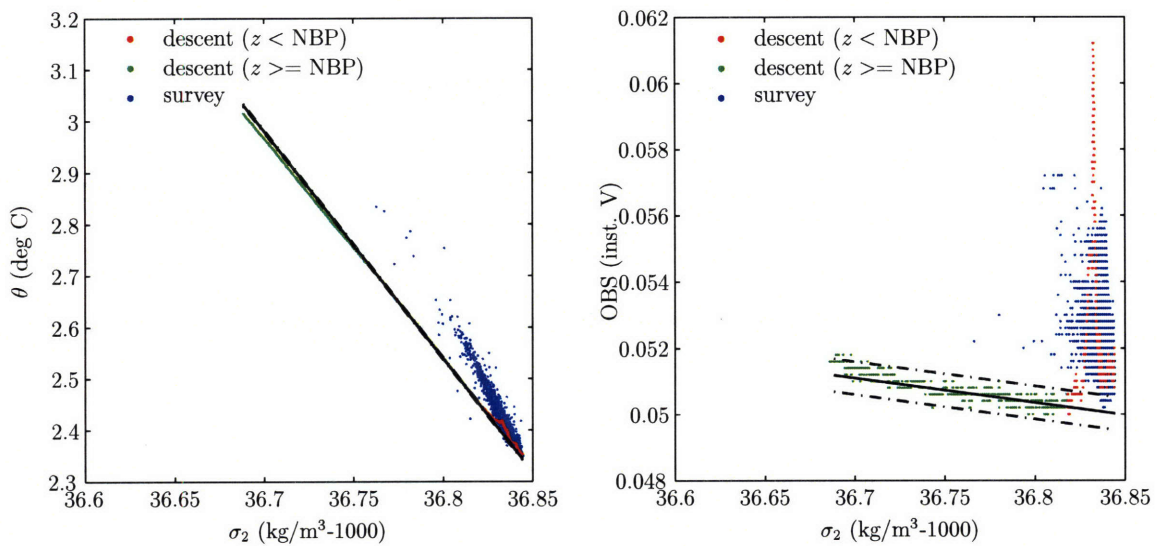


Figure F-27: Descent profiles vs. potential density, ABE-136.

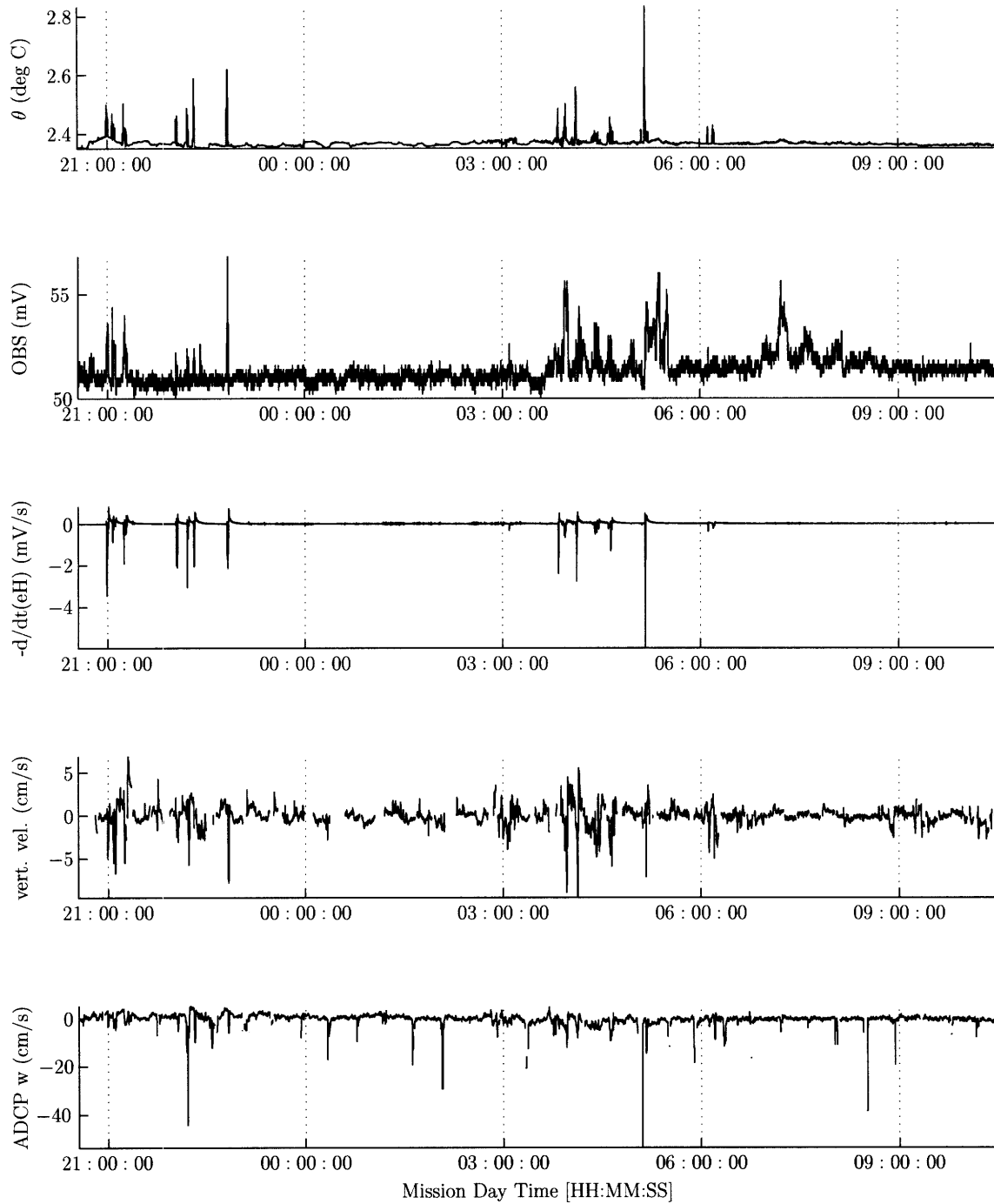


Figure F-28: Time series of hydrothermal tracers after pre-processing, ABE-136.

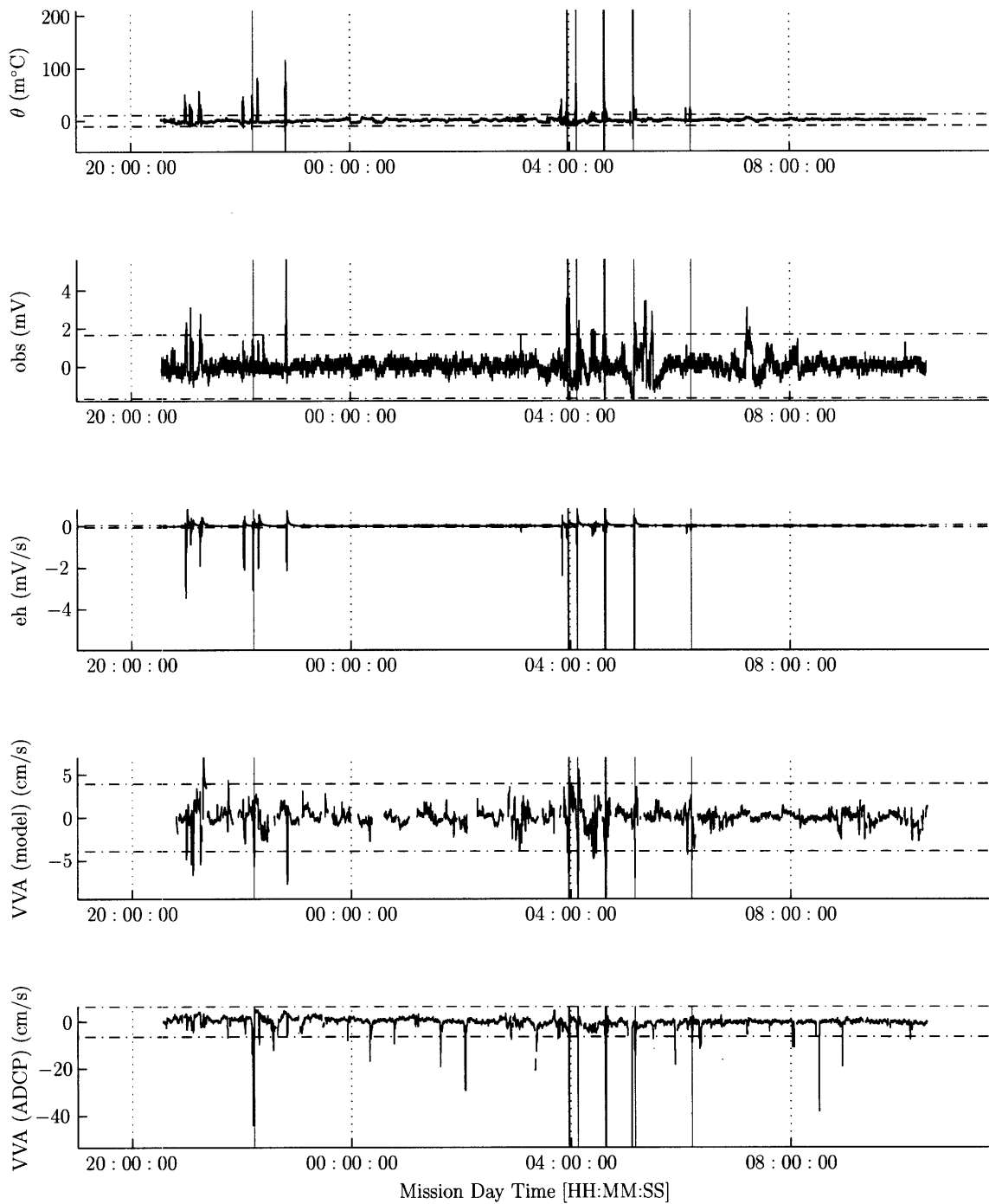


Figure F-29: Time series of hydrothermal tracers processed for BP detection, ABE-136. Intervals highlighted in gray indicate BP detection.

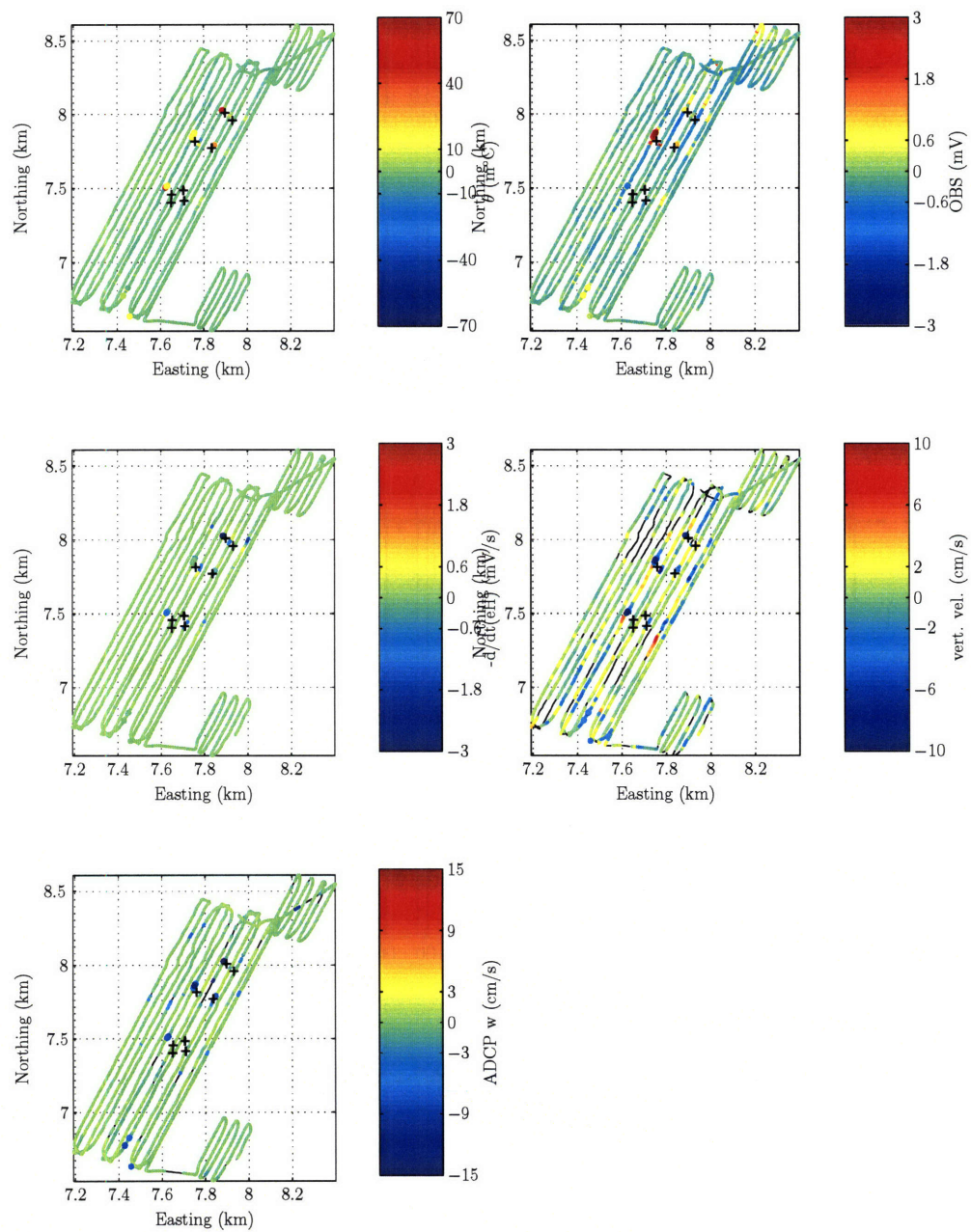


Figure F-30: Bird's eye view of anomaly intensity after processing for BP detection, ABE-136. Larger diameter dots indicate BP detection.

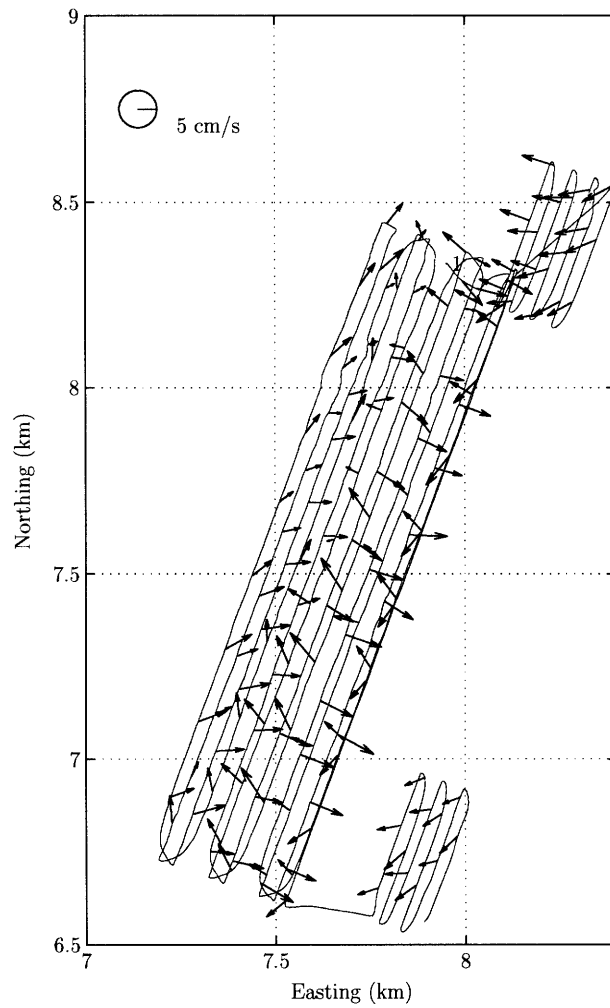


Figure F-31: Bird's eye view of ADCP-derived crossflow velocity, ABE-136 (5 min vertical average over depth interval 5 m to 25 m below vehicle). Numbers indicate the order of observation.

F.3 ELSC: Site-5

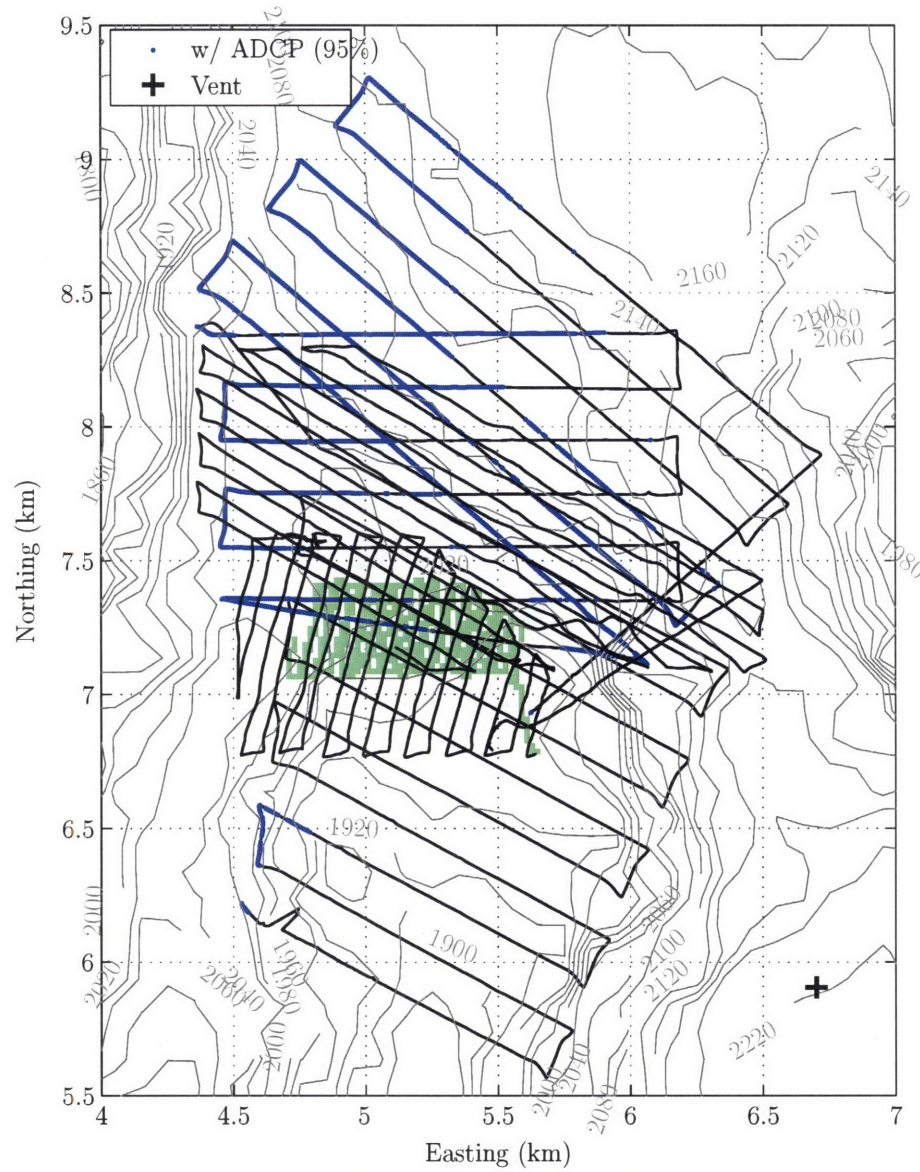


Figure F-32: Vent prospecting summary data, ELSC-5.

F.3.1 ABE-131

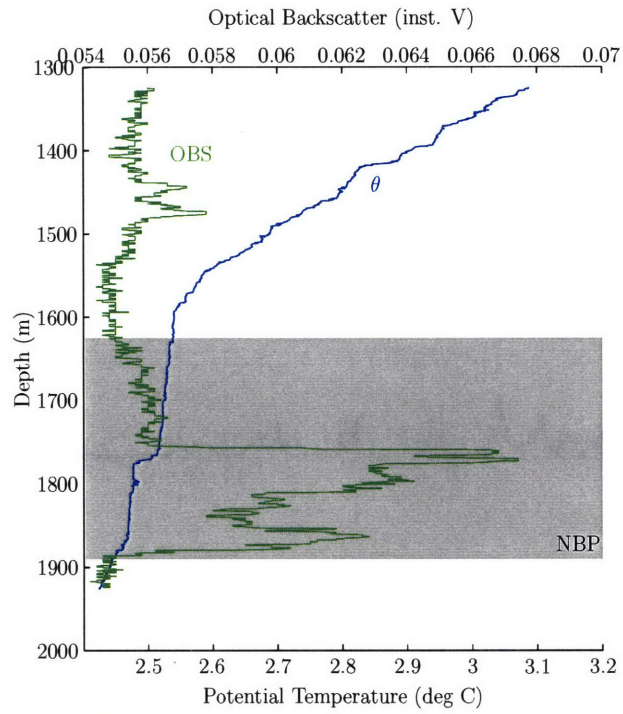


Figure F-33: Descent profiles vs. depth, ABE-131.

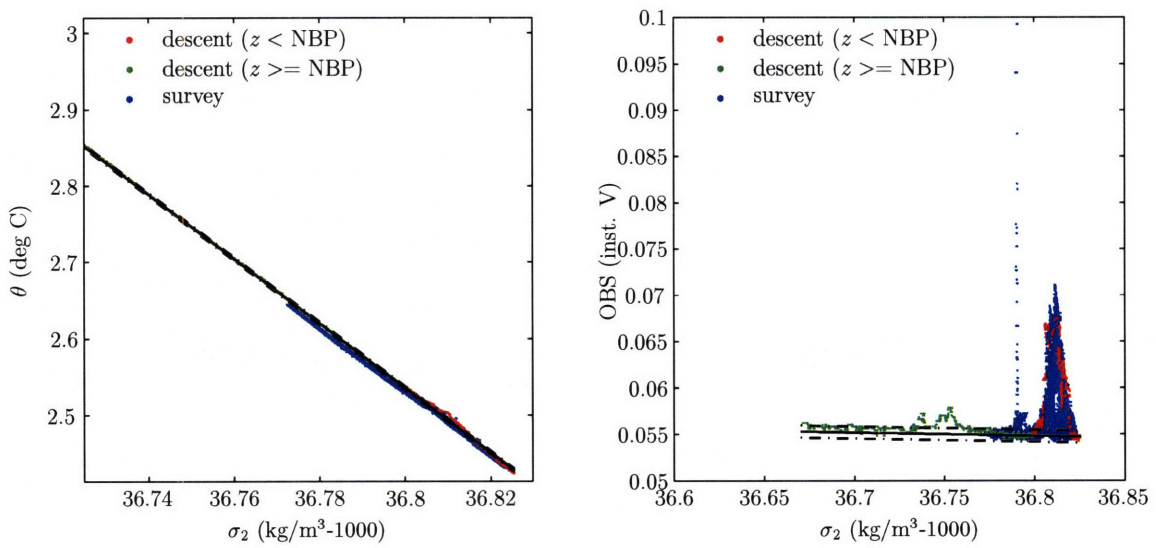


Figure F-34: Descent profiles vs. potential density, ABE-131.

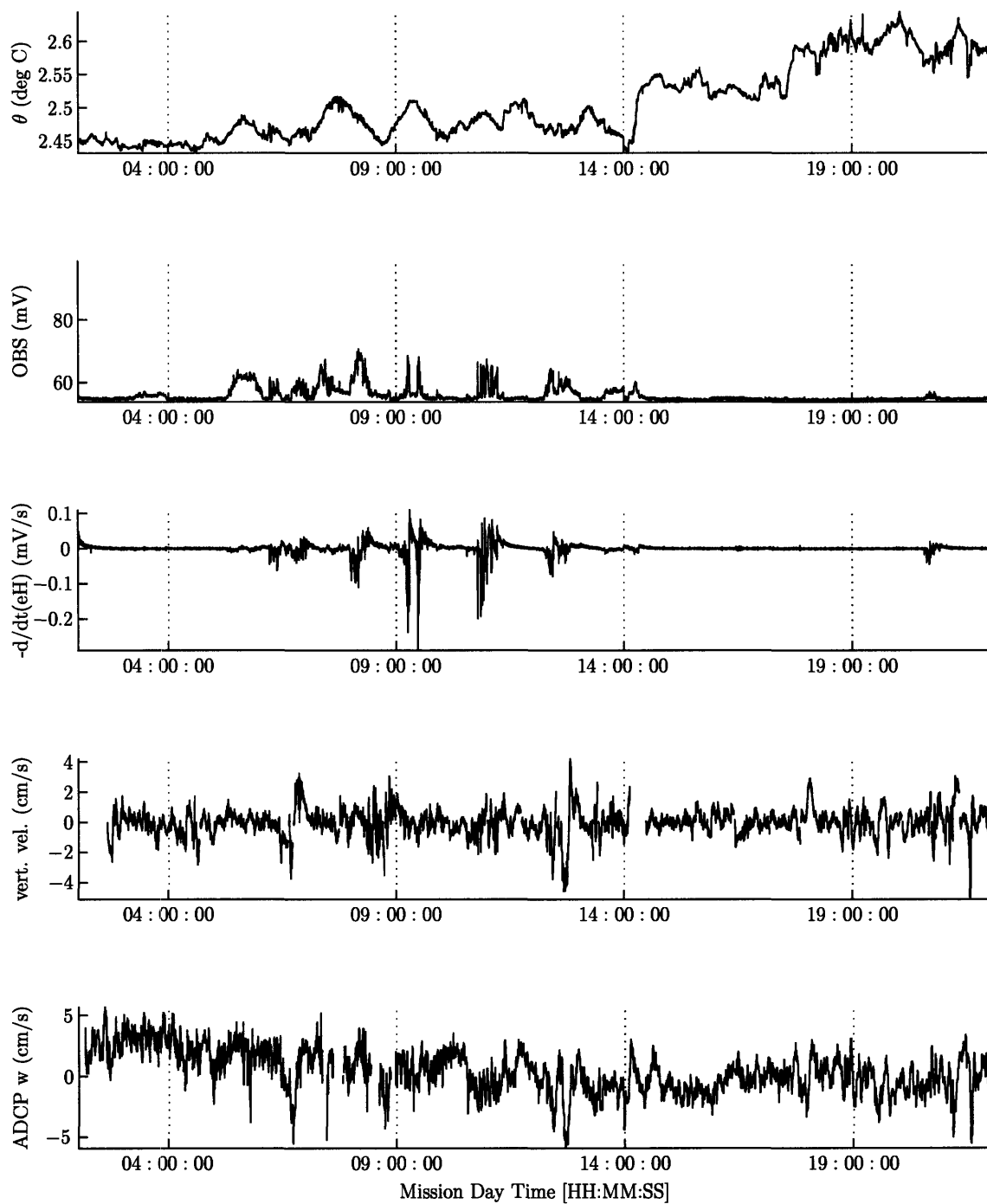


Figure F-35: Time series of hydrothermal tracers after pre-processing, ABE-131.

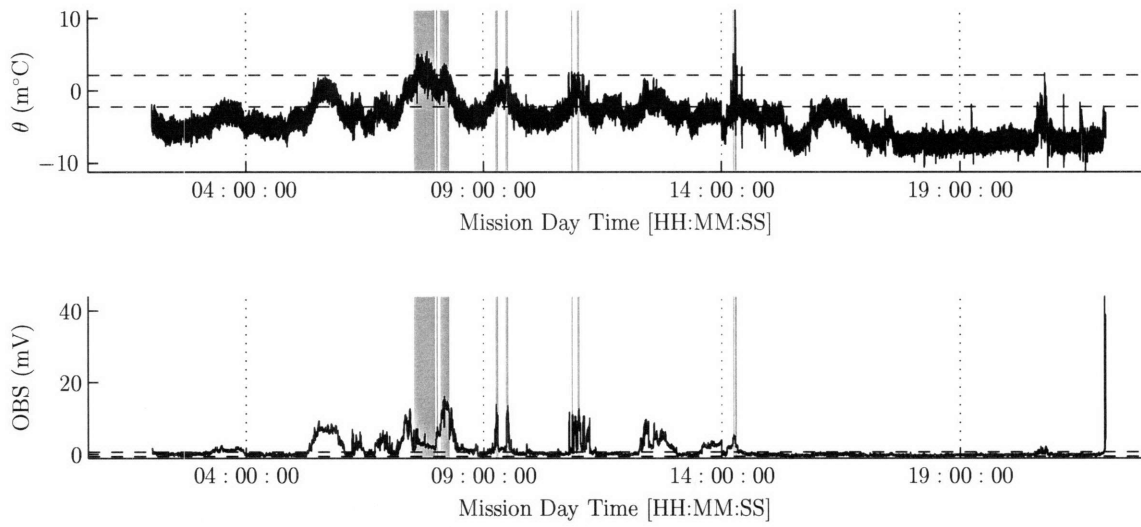


Figure F-36: Time series of hydrothermal tracers processed for NBP detection, ABE-131. Intervals highlighted in gray indicate NBP detection.

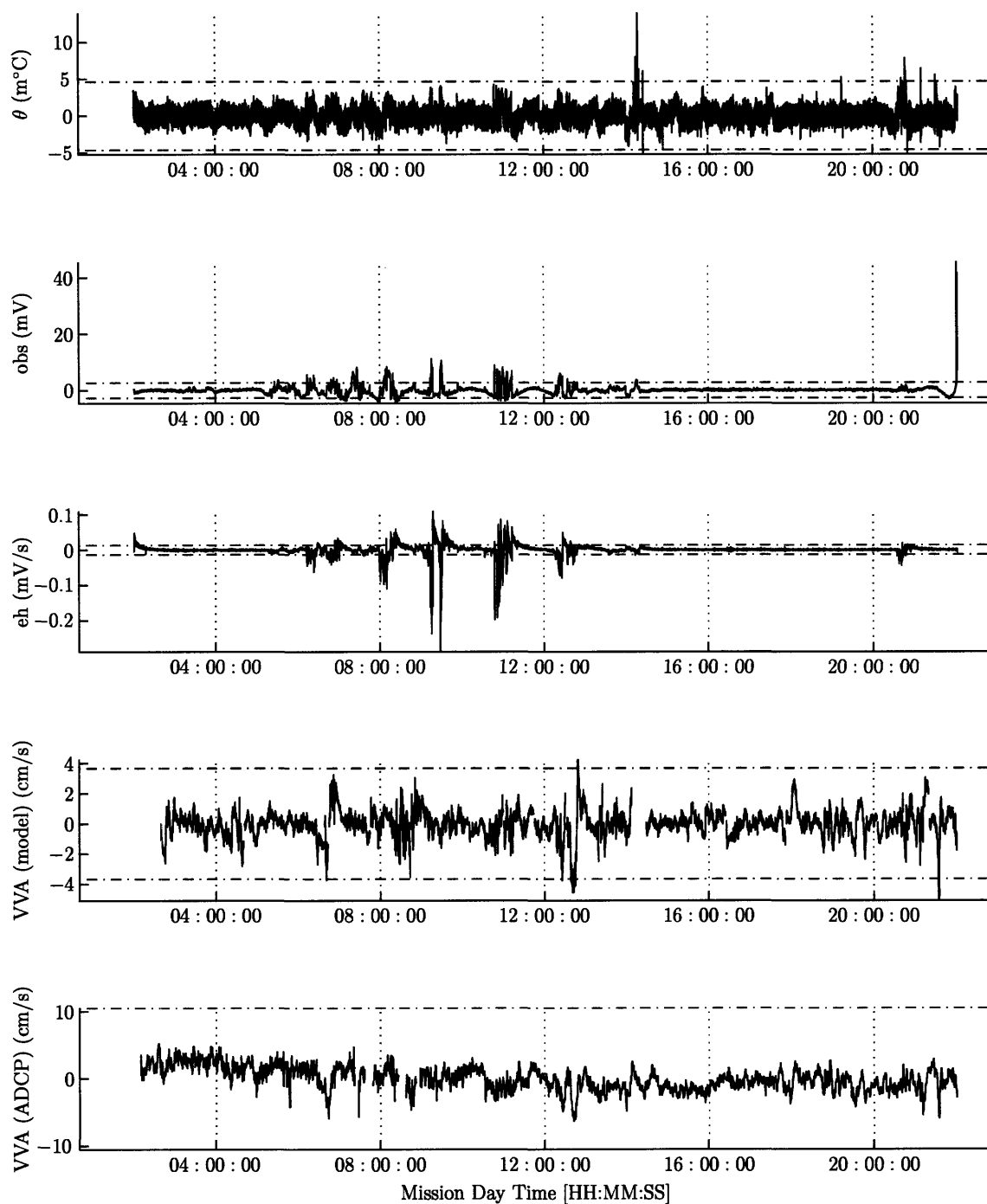


Figure F-37: Time series of hydrothermal tracers processed for BP detection, ABE-131. Intervals highlighted in gray indicate BP detection.

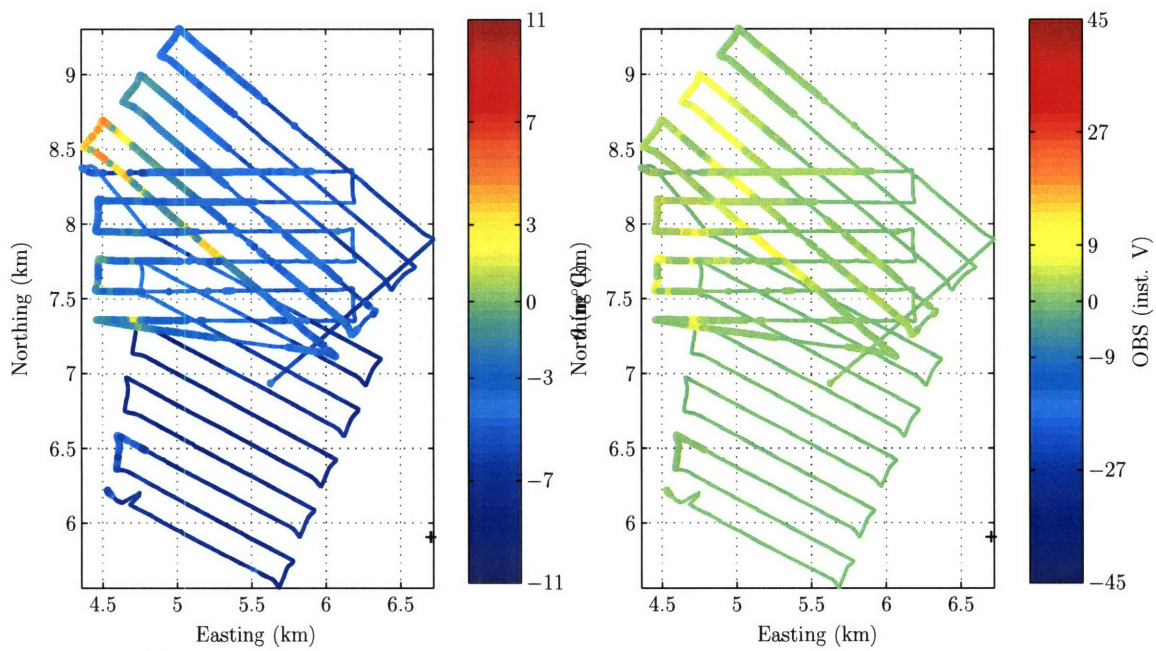


Figure F-38: Bird's eye view of anomaly intensity after processing for NBP detection, ABE-131. Larger diameter dots indicate NBP detection.

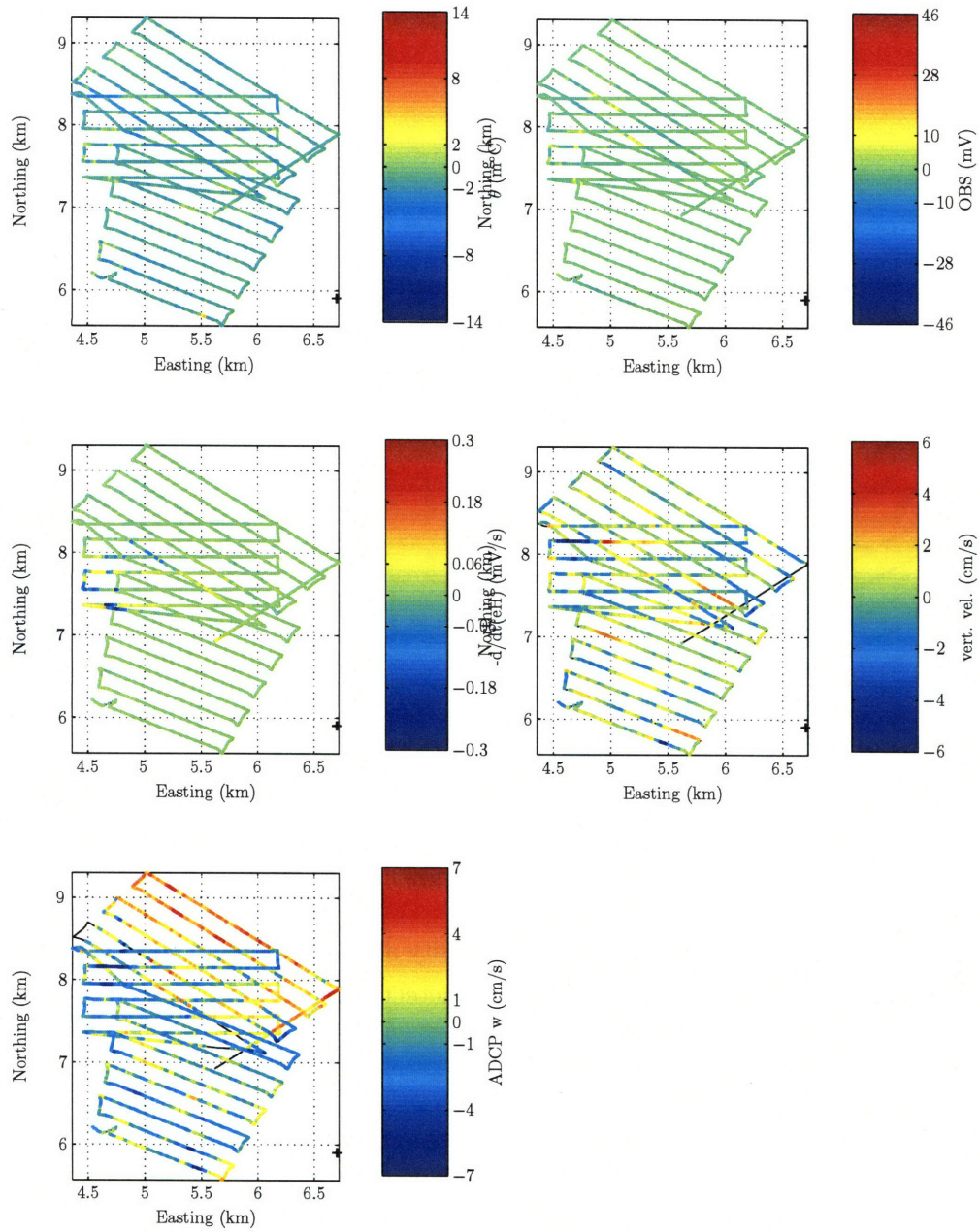


Figure F-39: Bird's eye view of anomaly intensity after processing for BP detection, ABE-131. Larger diameter dots indicate BP detection.

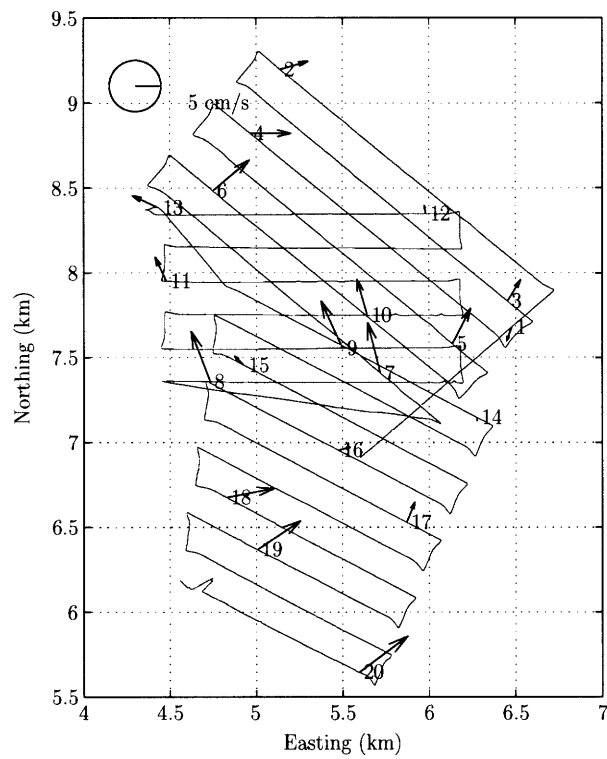


Figure F-40: Bird's eye view of ADCP-derived crossflow velocity, ABE-131 (1 h vertical average over depth interval 5 m to 25 m below vehicle). Numbers indicate the order of observation.

F.3.2 ABE-133

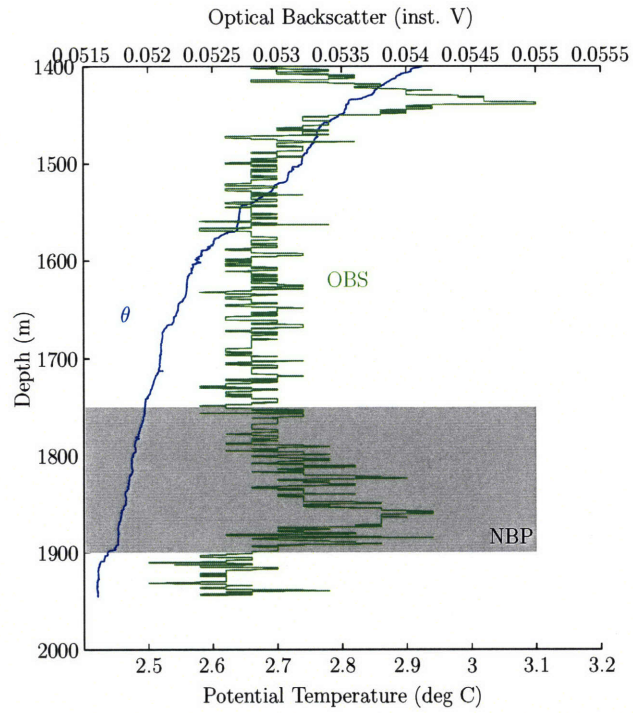


Figure F-41: Descent profiles vs. depth, ABE-133.

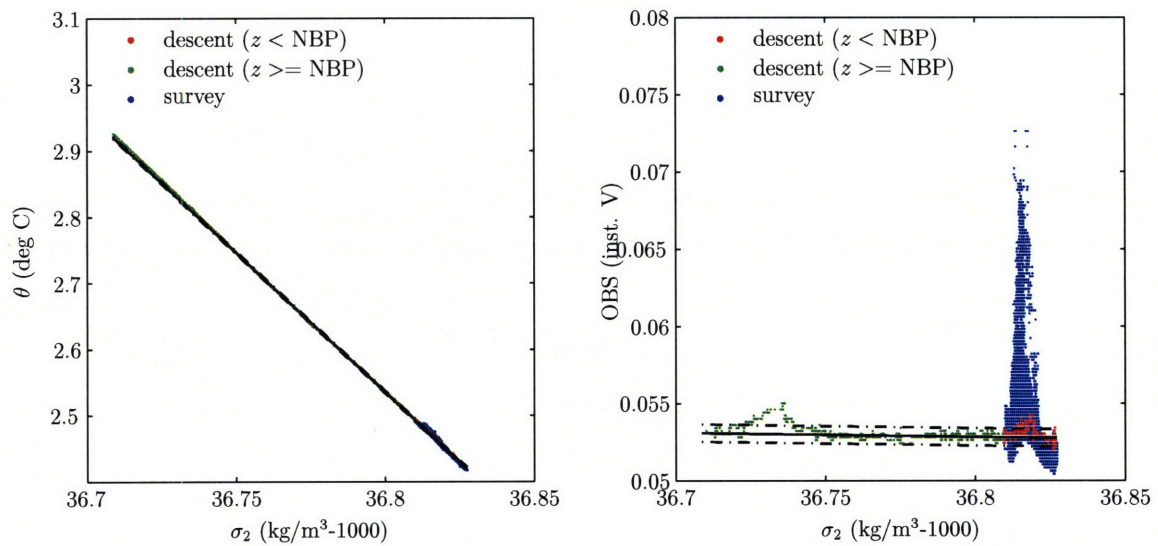


Figure F-42: Descent profiles vs. potential density, ABE-133.

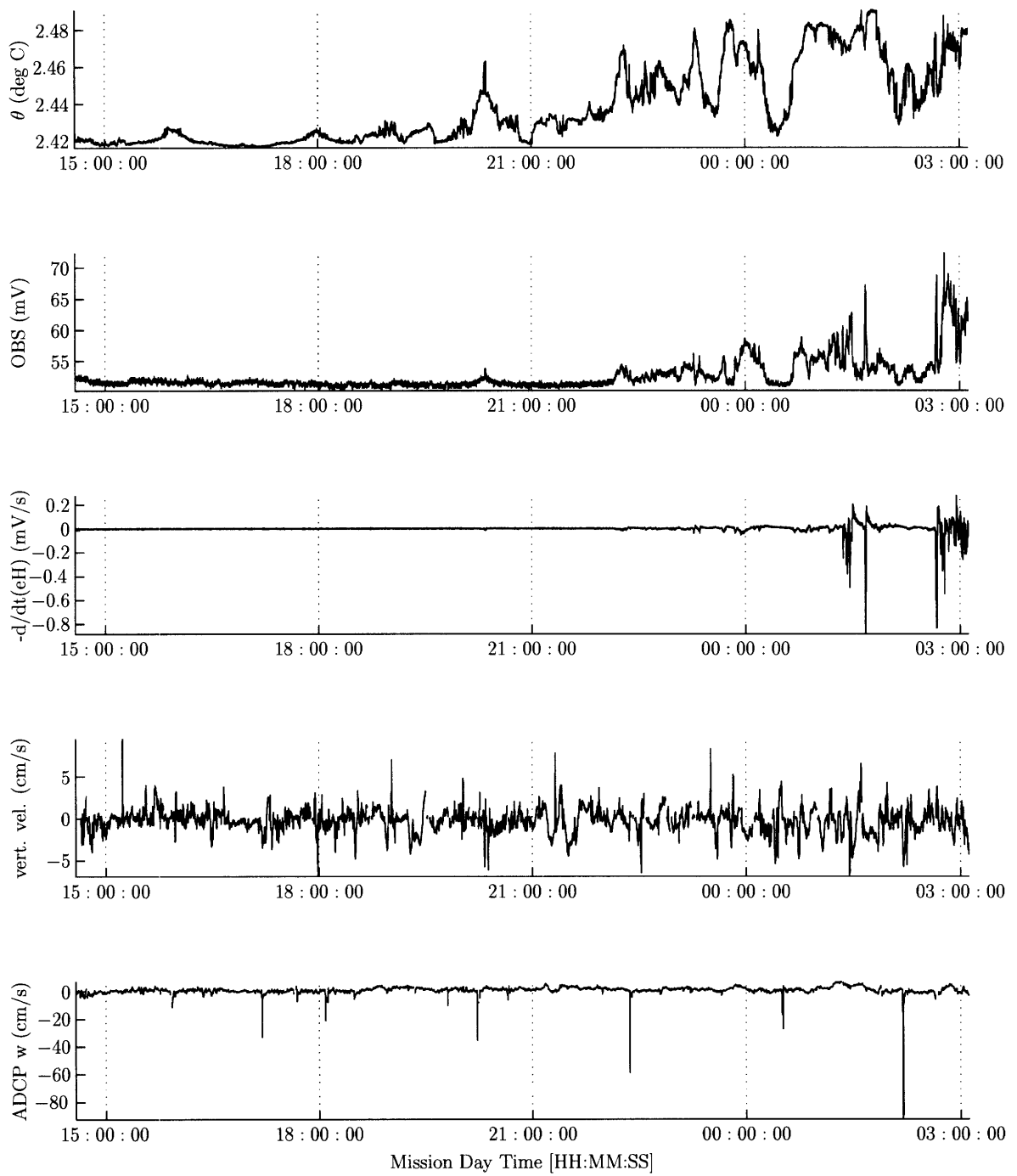


Figure F-43: Time series of hydrothermal tracers after pre-processing, ABE-133.

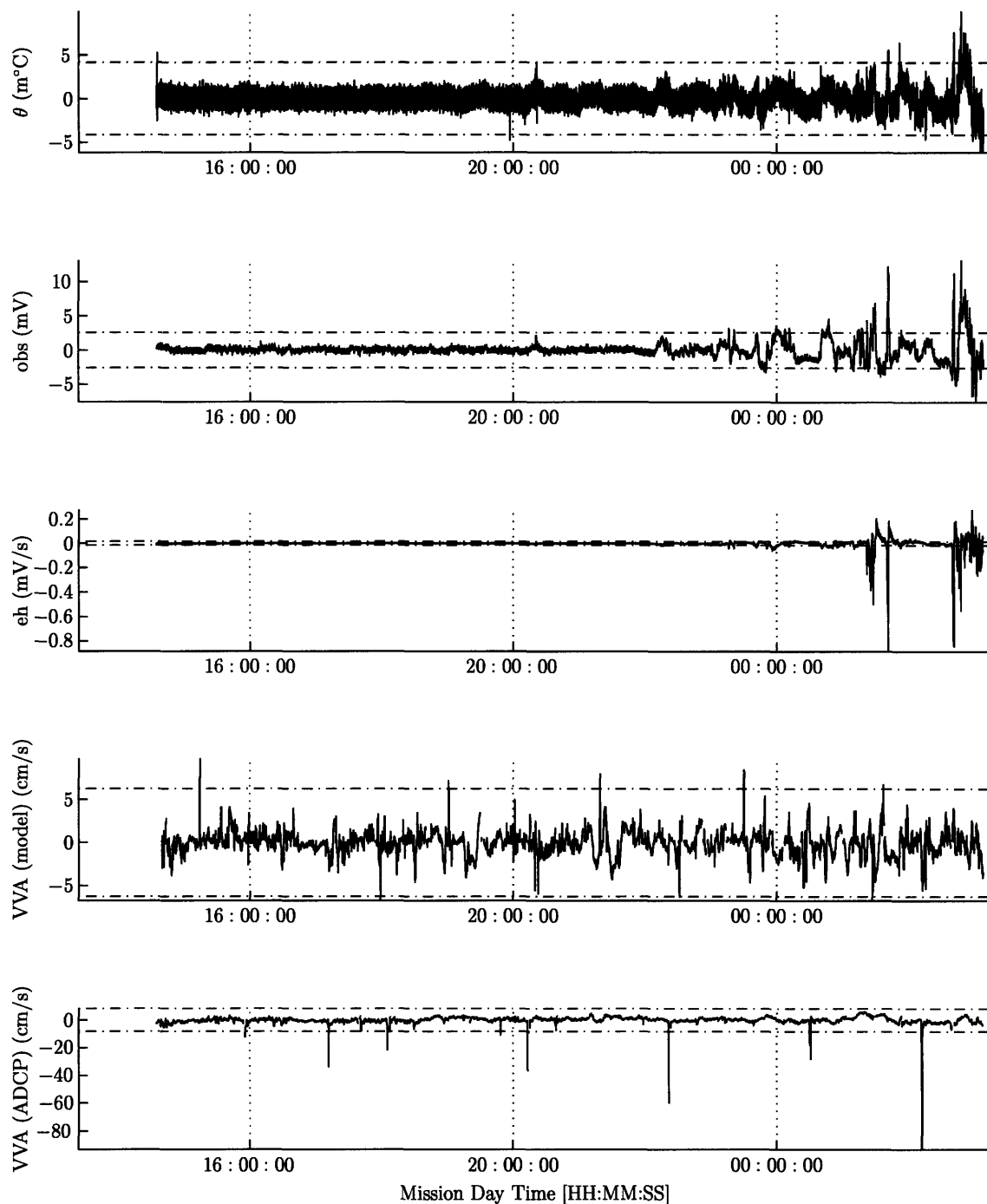


Figure F-44: Time series of hydrothermal tracers processed for BP detection, ABE-133. Intervals highlighted in gray indicate BP detection.

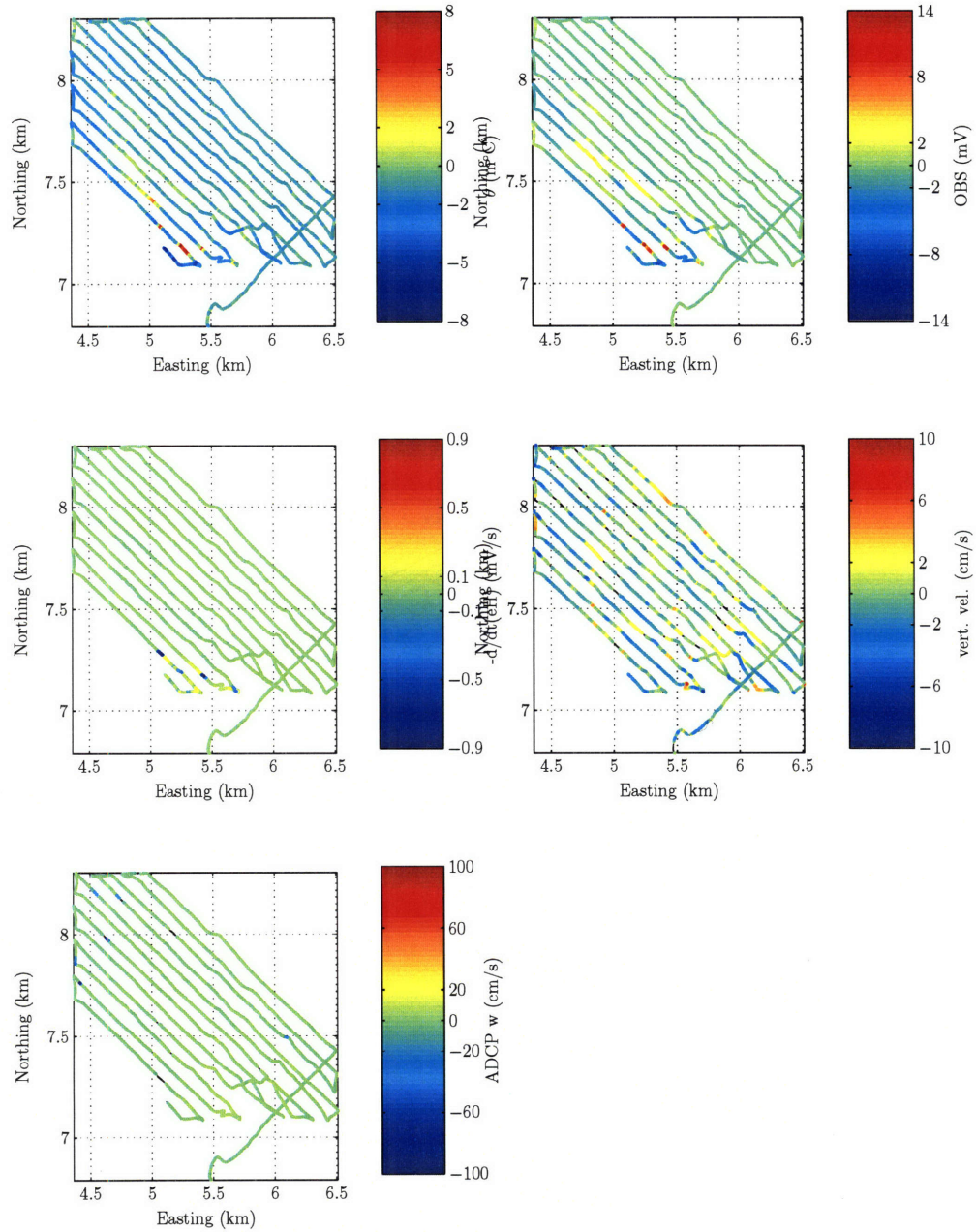


Figure F-45: Bird's eye view of anomaly intensity after processing for BP detection, ABE-133. Larger diameter dots indicate BP detection.

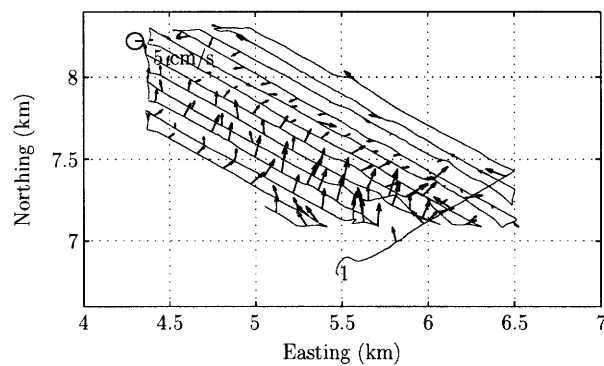


Figure F-46: Bird's eye view of ADCP-derived crossflow velocity, ABE-133 (5 min vertical average over depth interval 5 m to 25 m below vehicle). Numbers indicate the order of observation.

F.3.3 ABE-134

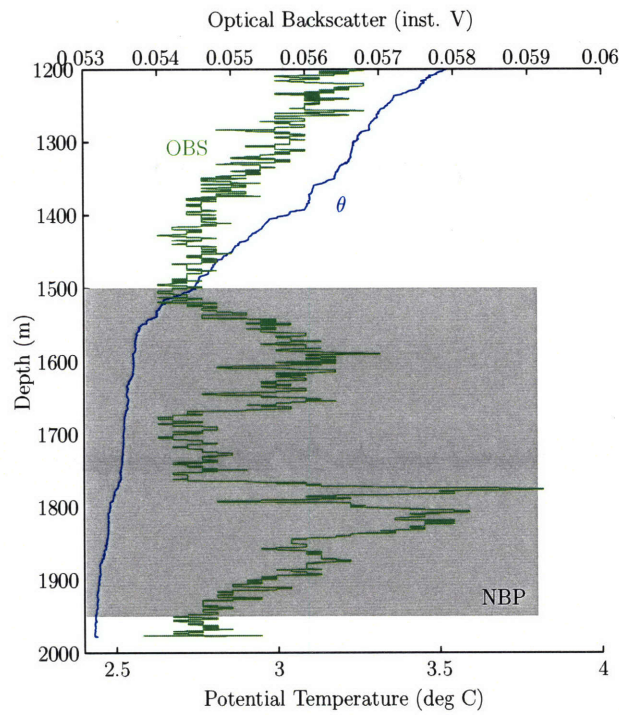


Figure F-47: Descent profiles vs. depth, ABE-134.

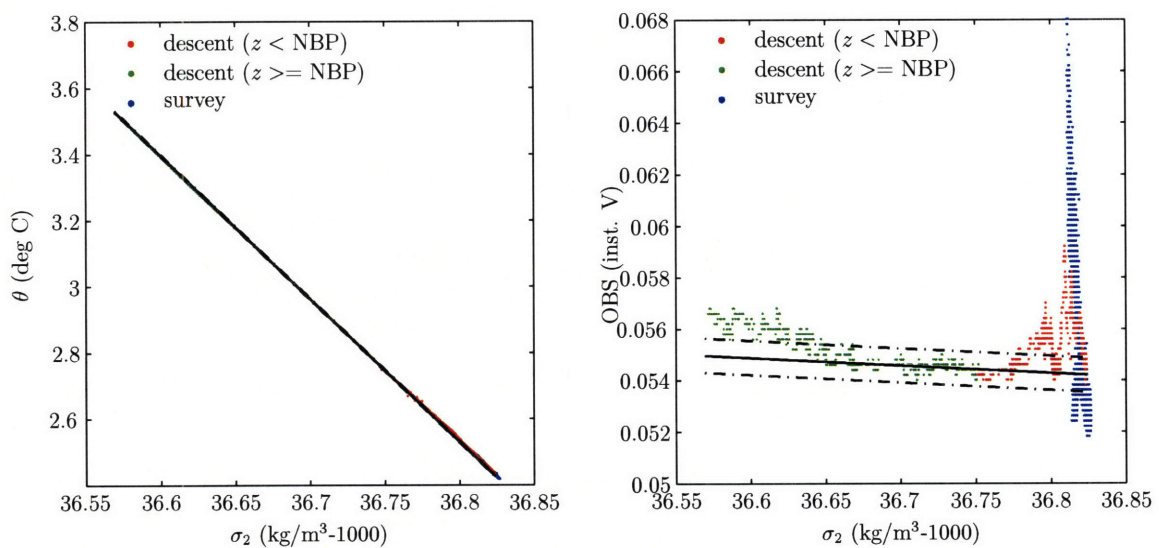


Figure F-48: Descent profiles vs. potential density, ABE-134.

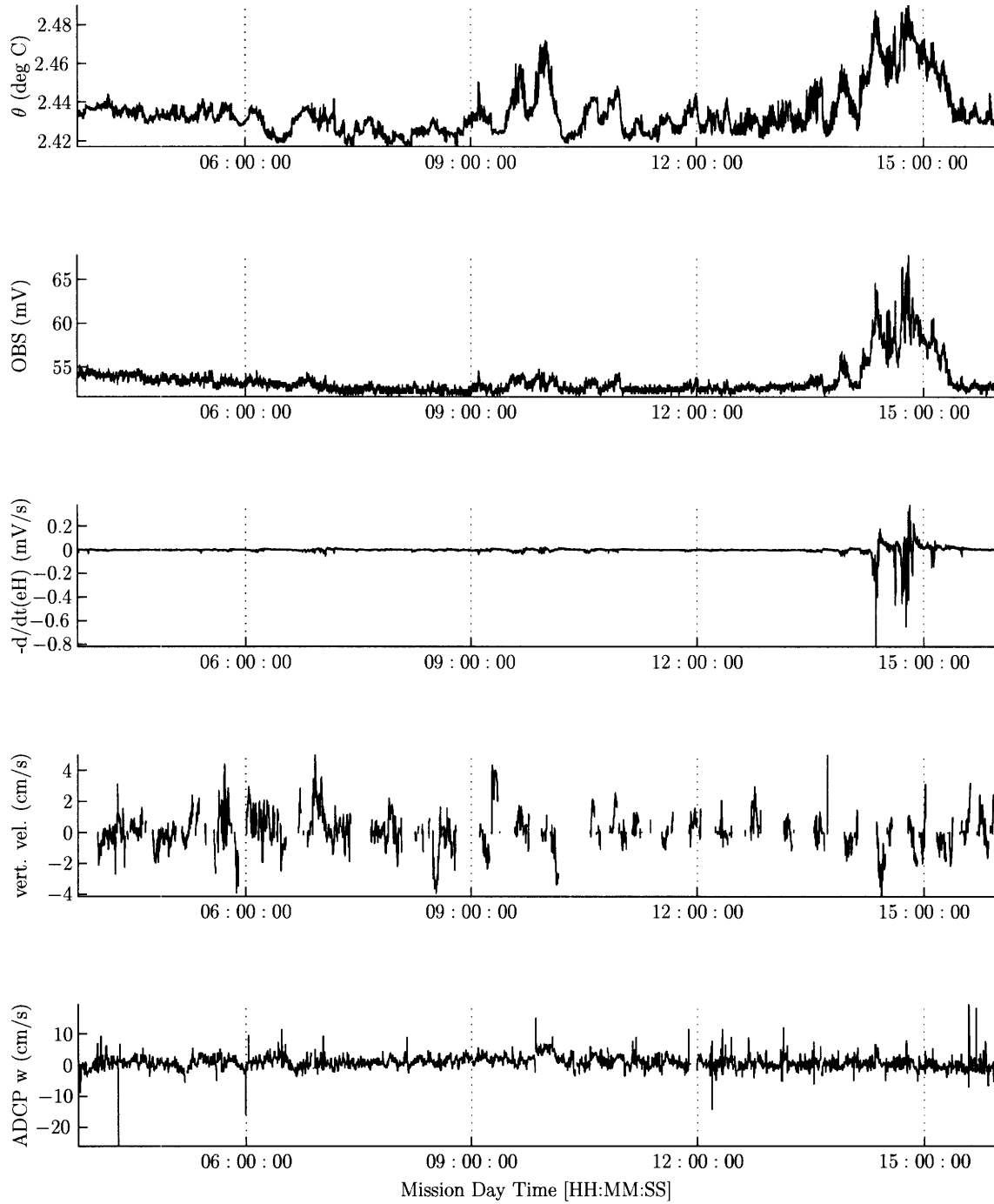


Figure F-49: Time series of hydrothermal tracers after pre-processing, ABE-134.

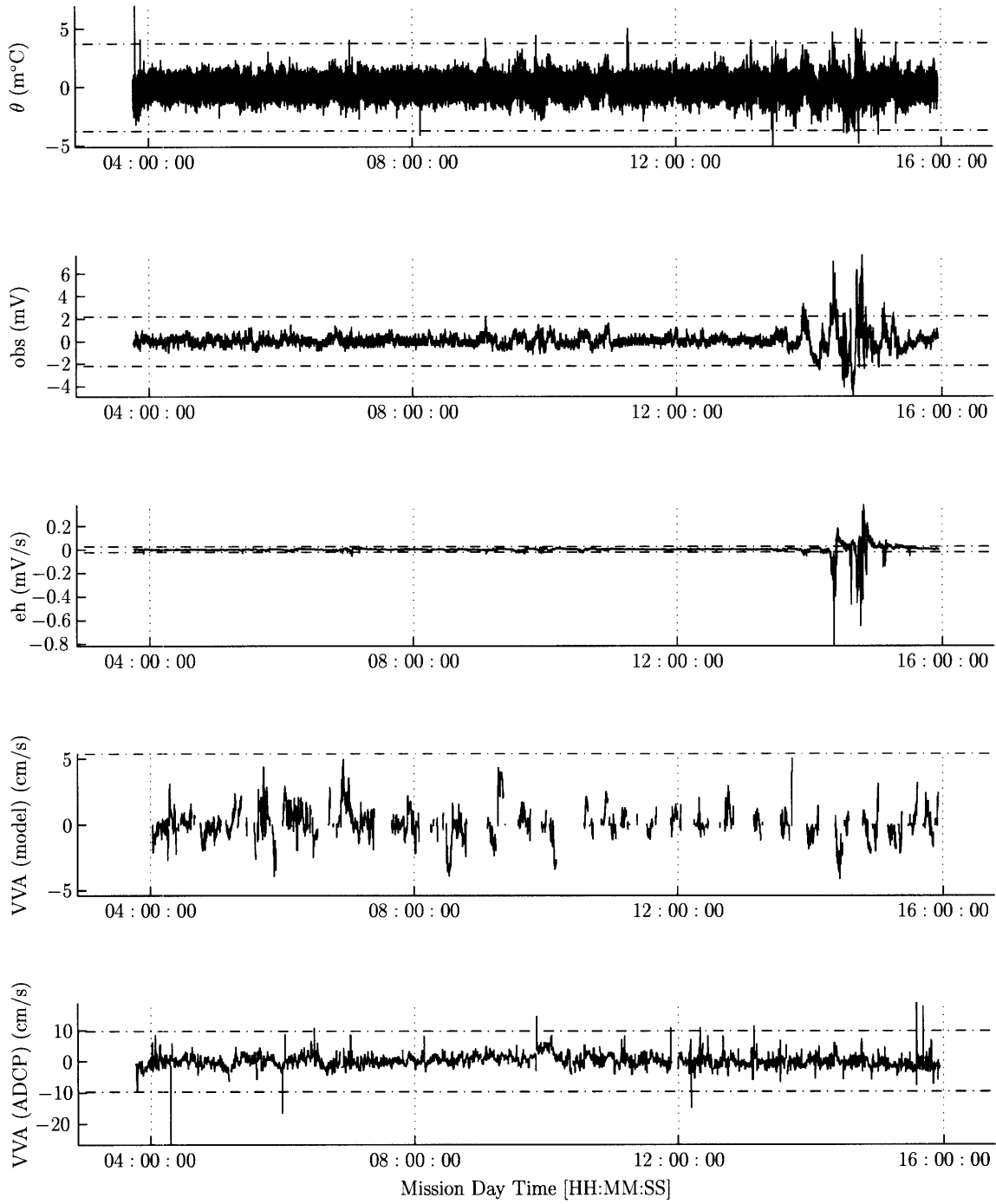


Figure F-50: Time series of hydrothermal tracers processed for BP detection, ABE-134. Intervals highlighted in gray indicate BP detection.

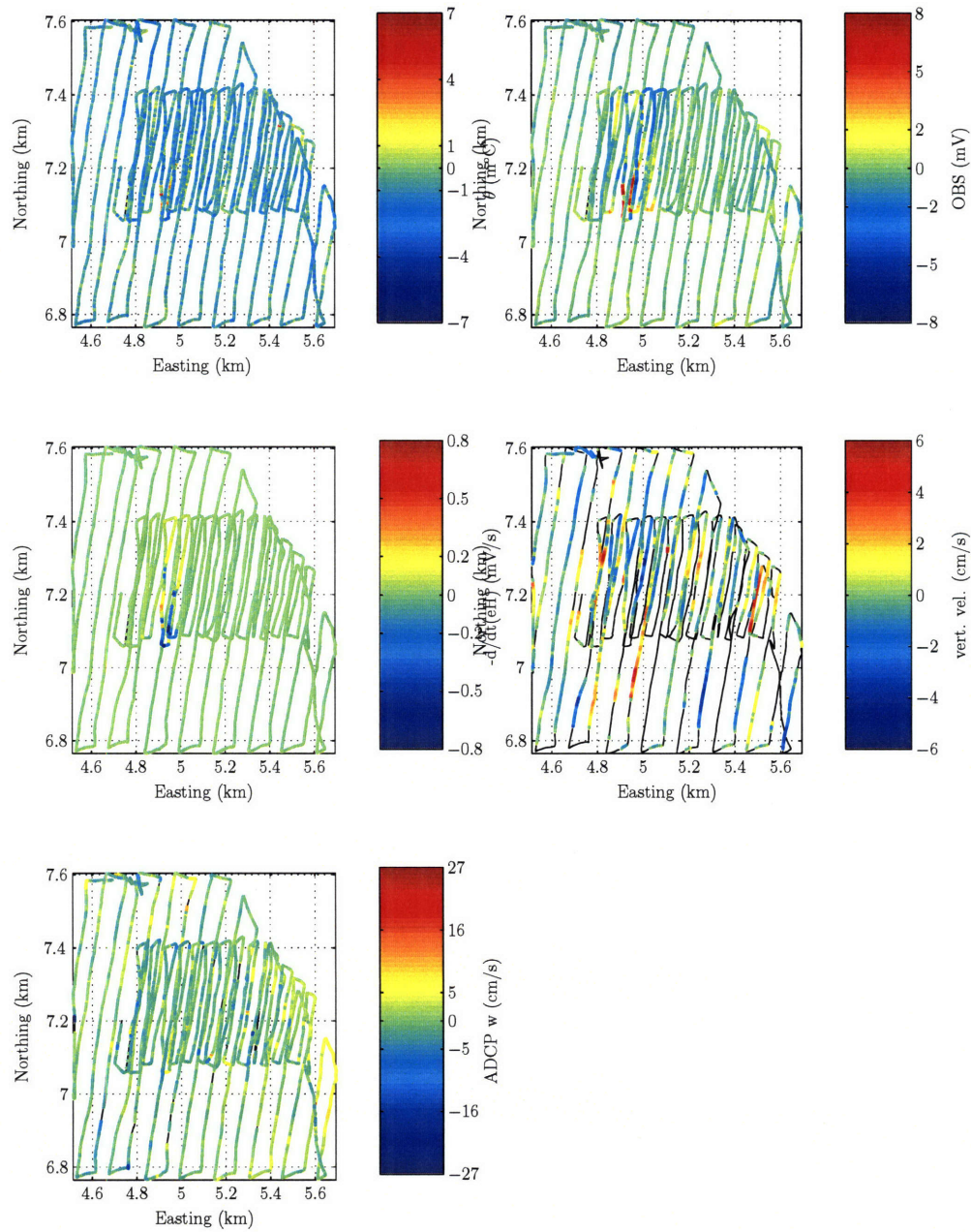


Figure F-51: Bird's eye view of anomaly intensity after processing for BP detection, ABE-134. Larger diameter dots indicate BP detection.

F.4 SMAR

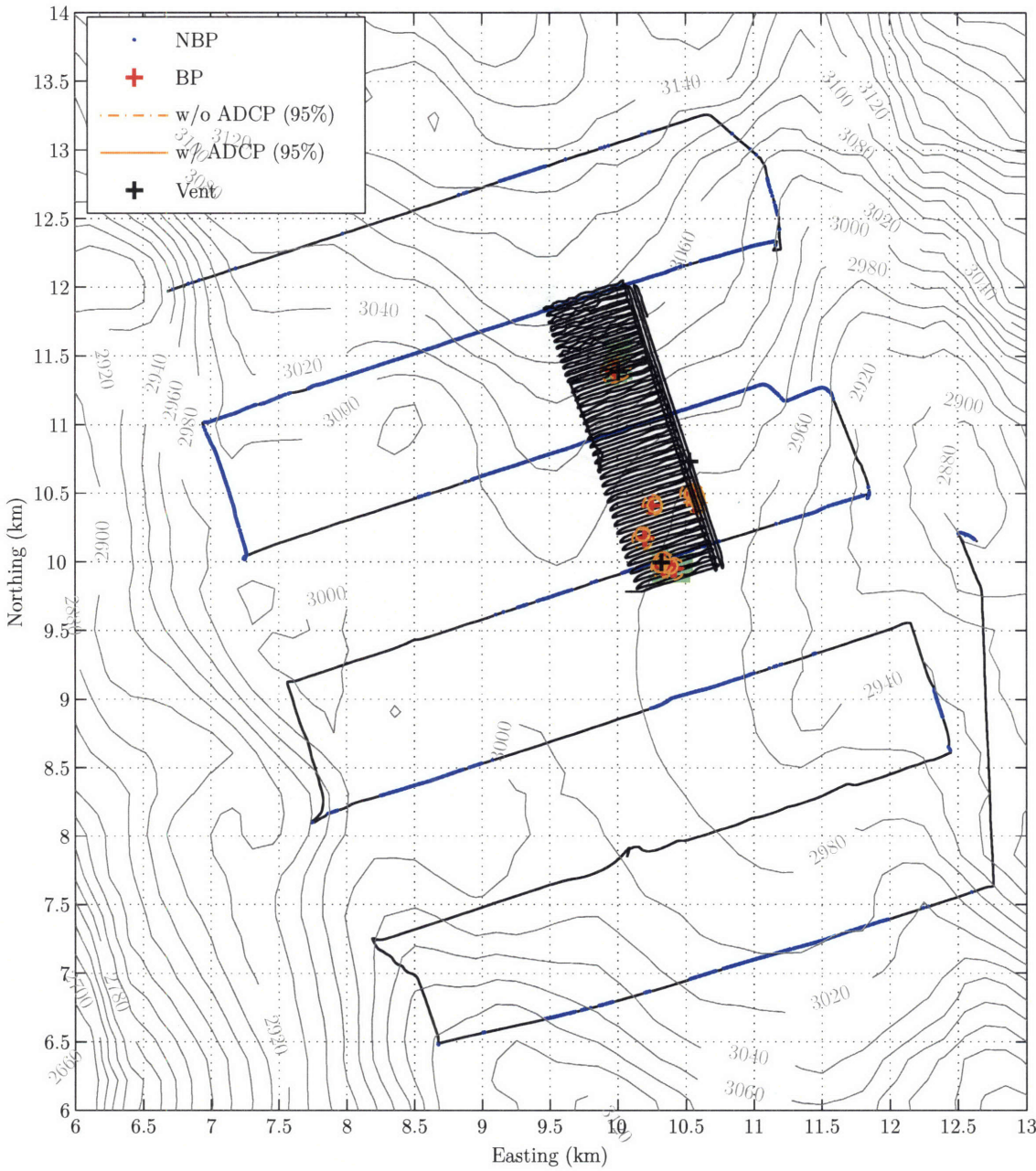


Figure F-52: Vent prospecting summary data, SMAR.

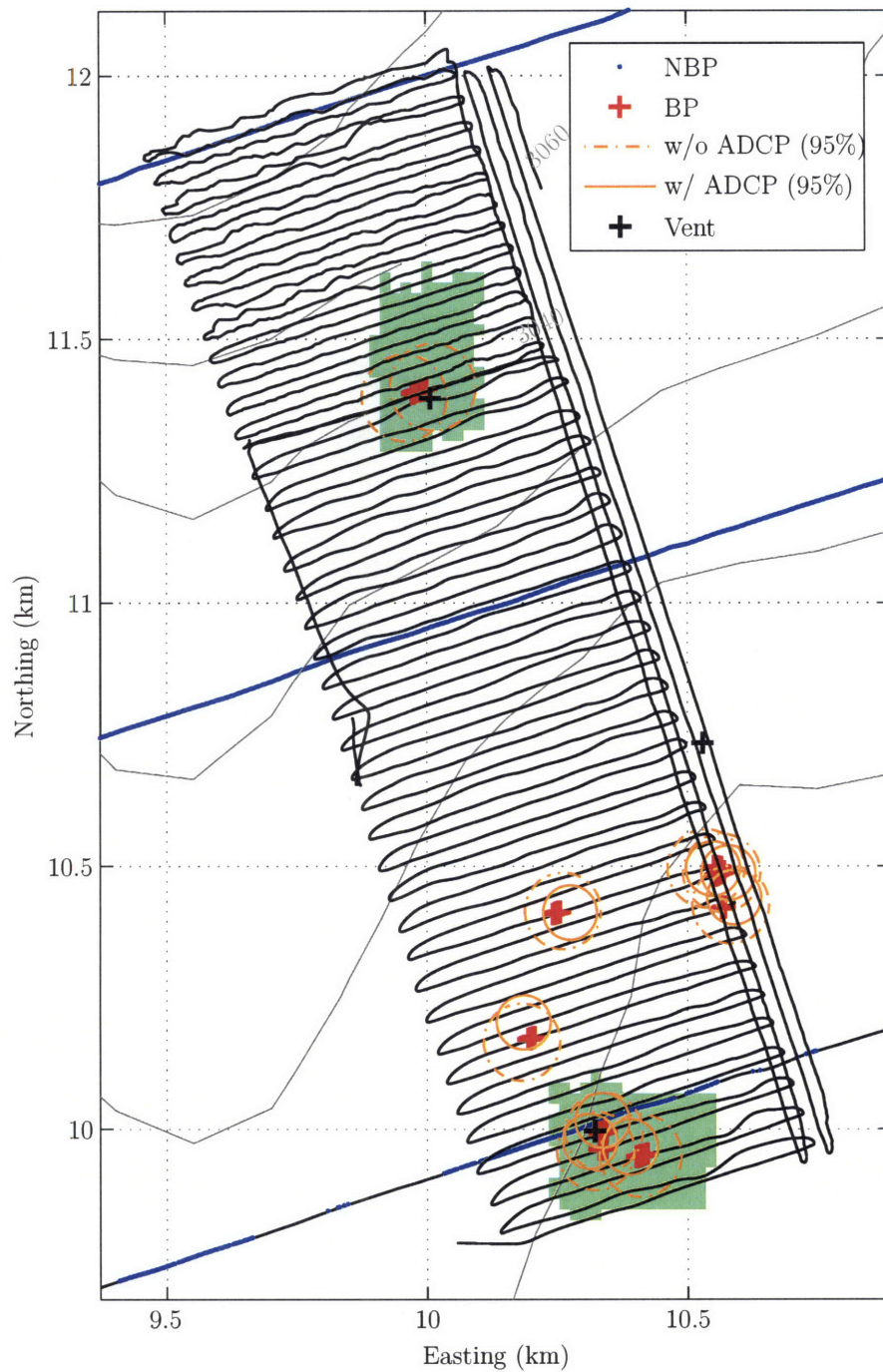


Figure F-53: Closeup of vent prospecting summary data, SMAR.

F.4.1 ABE-150

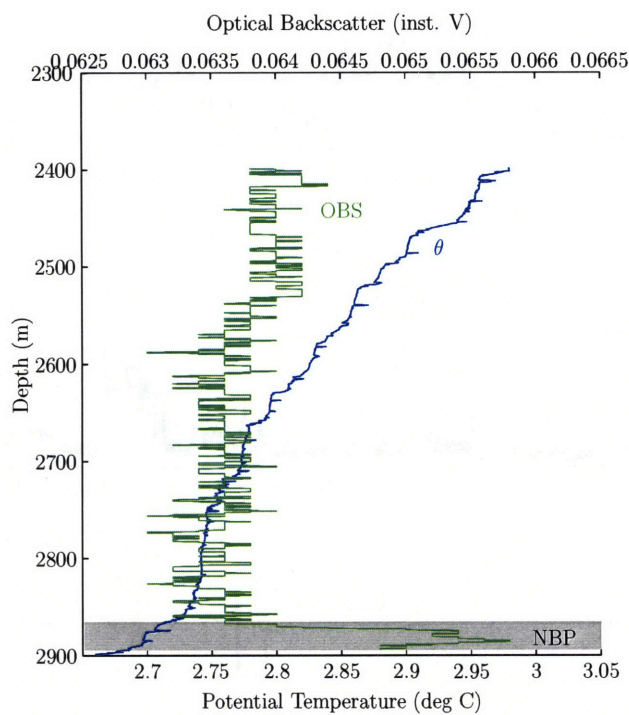


Figure F-54: Descent profiles vs. depth, ABE-150.

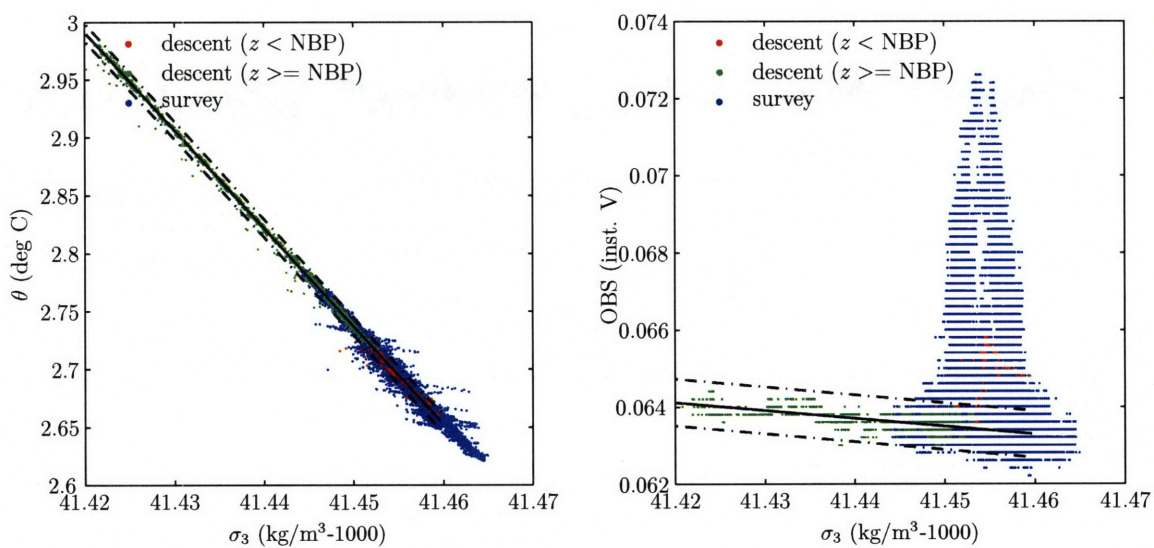


Figure F-55: Descent profiles vs. potential density, ABE-150.

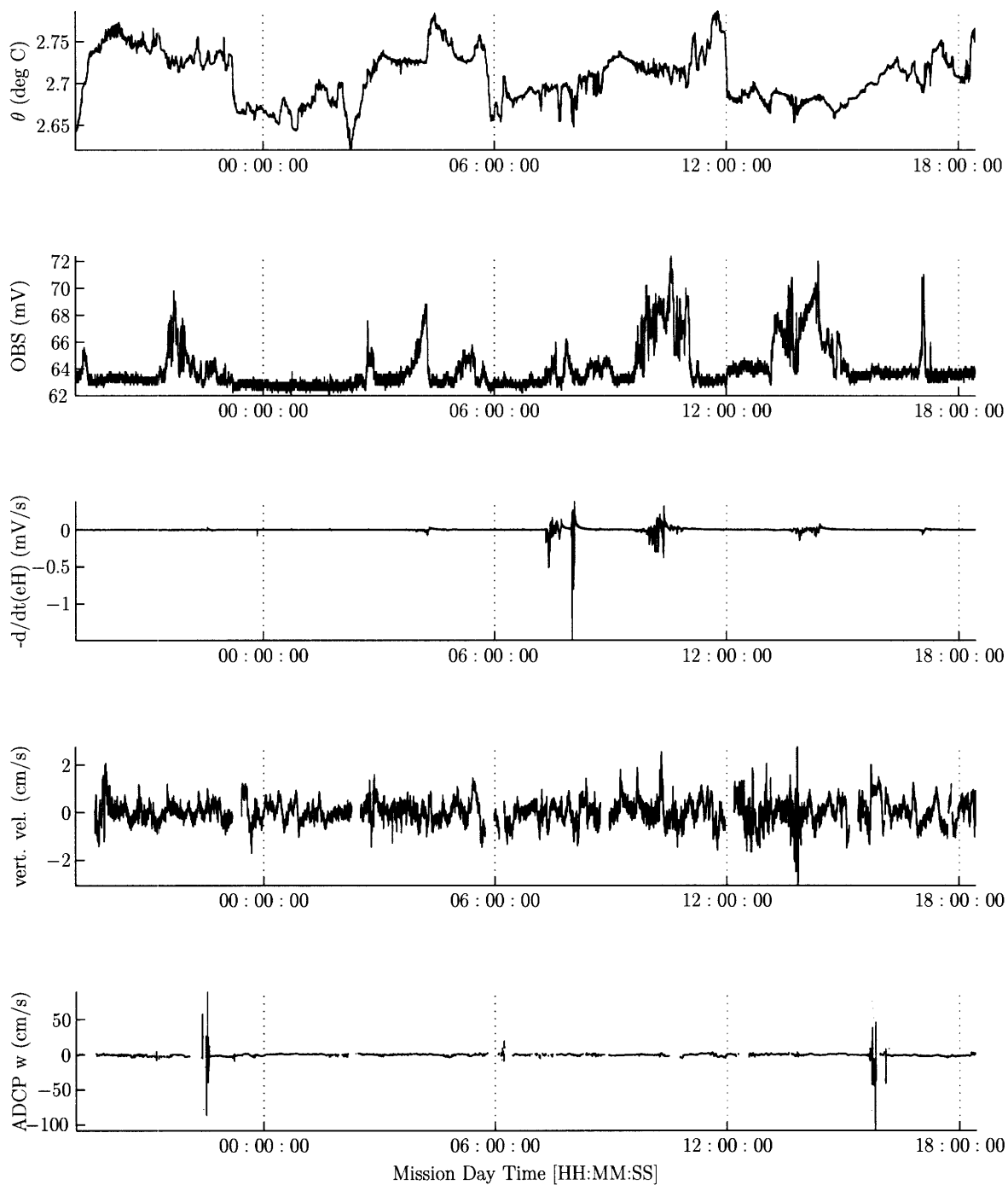


Figure F-56: Time series of hydrothermal tracers after pre-processing, ABE-150.

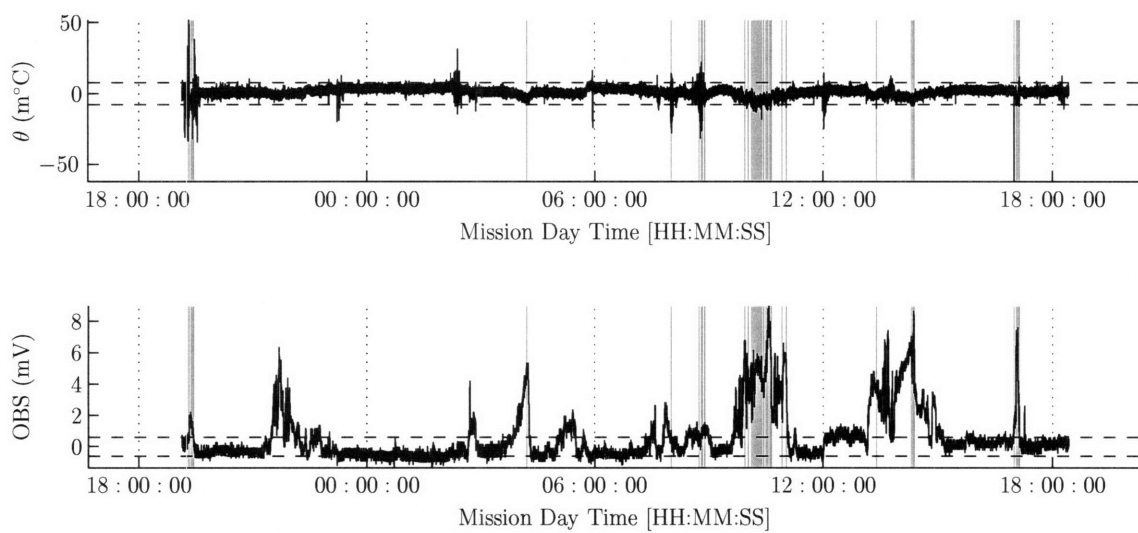


Figure F-57: Time series of hydrothermal tracers processed for NBP detection, ABE-150. Intervals highlighted in gray indicate NBP detection.

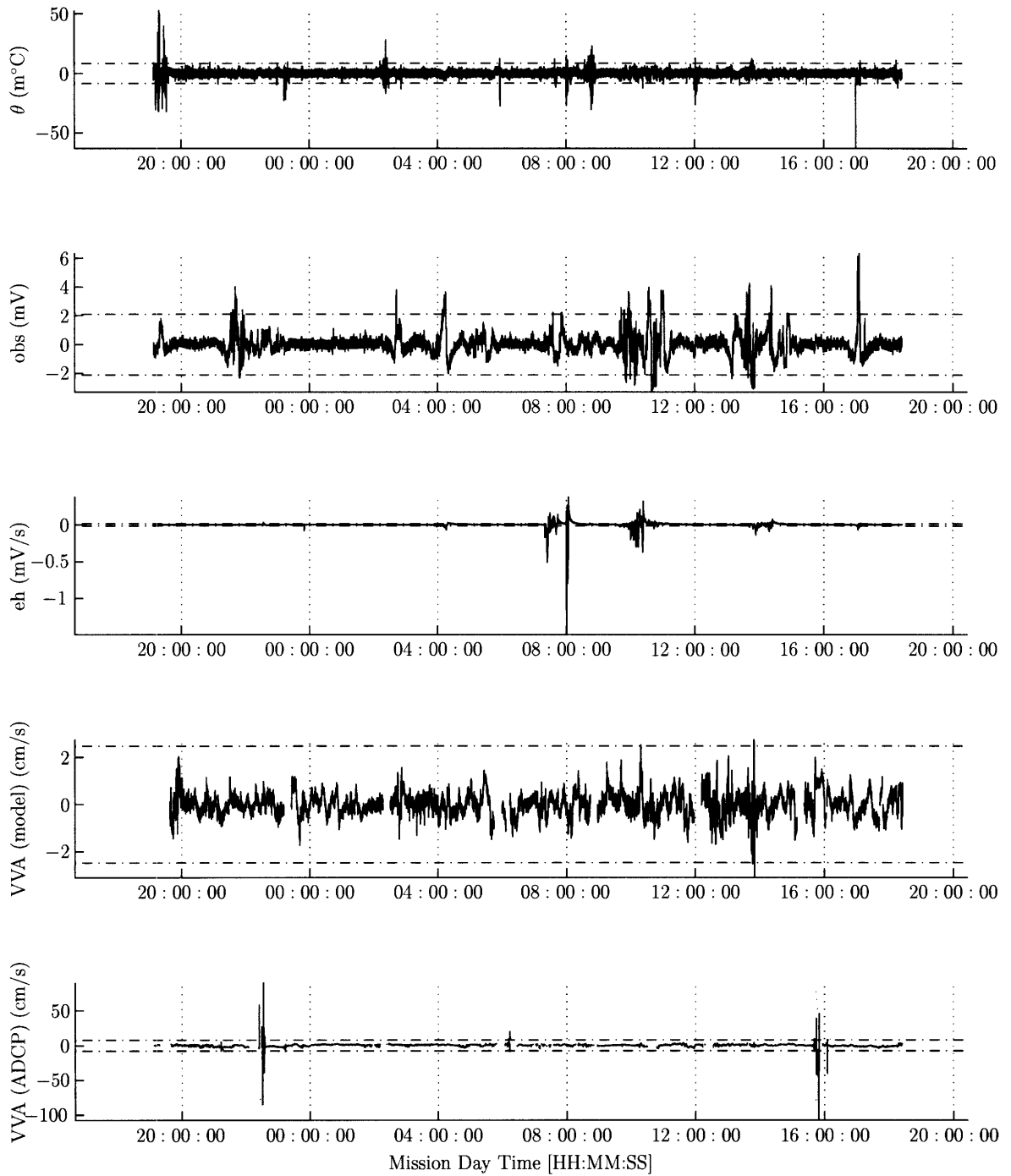


Figure F-58: Time series of hydrothermal tracers processed for BP detection, ABE-150. Intervals highlighted in gray indicate BP detection.

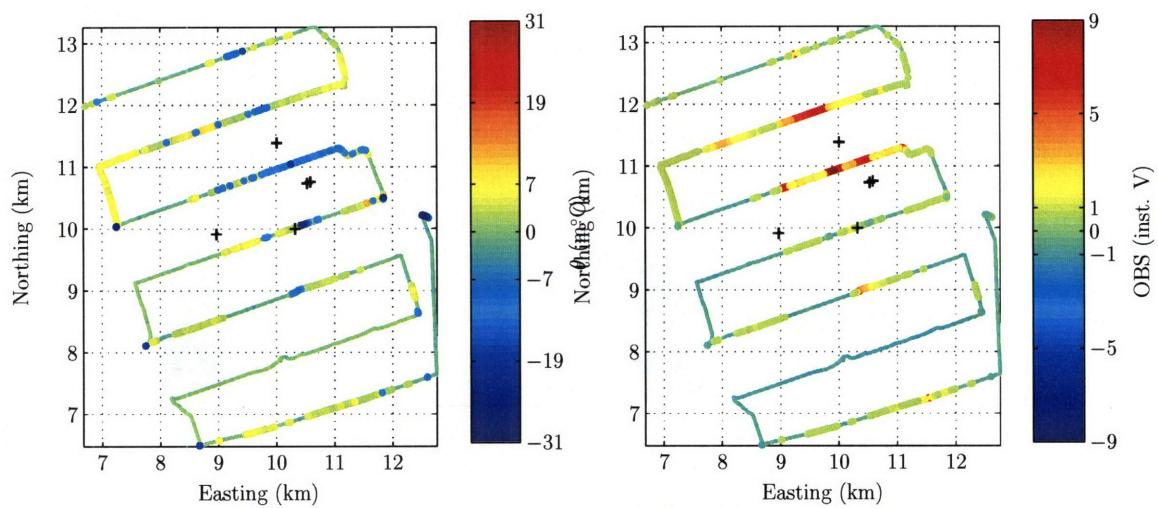


Figure F-59: Bird's eye view of anomaly intensity after processing for NBP detection, ABE-150. Larger diameter dots indicate NBP detection.

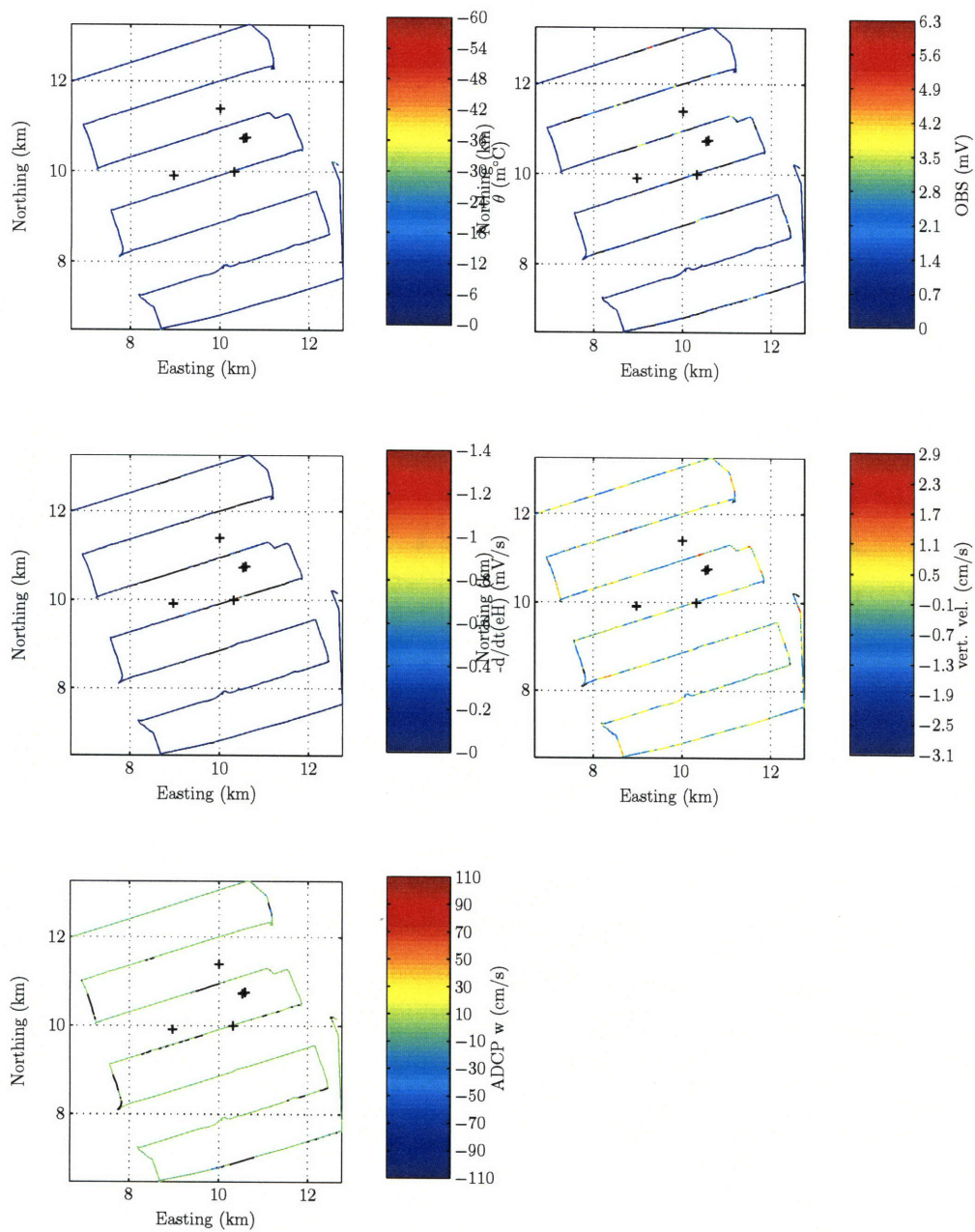


Figure F-60: Bird's eye view of anomaly intensity after processing for BP detection, ABE-150. Larger diameter dots indicate BP detection.

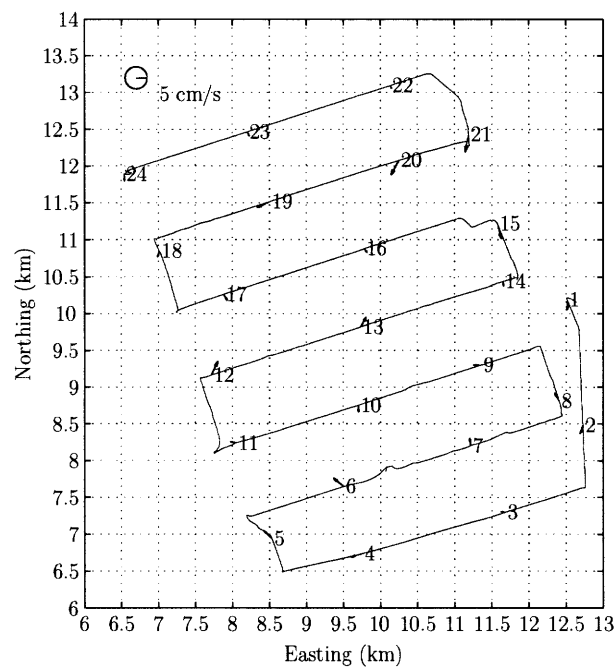


Figure F-61: Bird's eye view of ADCP-derived crossflow velocity, ABE-150 (1 h vertical average over depth interval 5 m to 25 m below vehicle). Numbers indicate the order of observation.

F.4.2 ABE-151

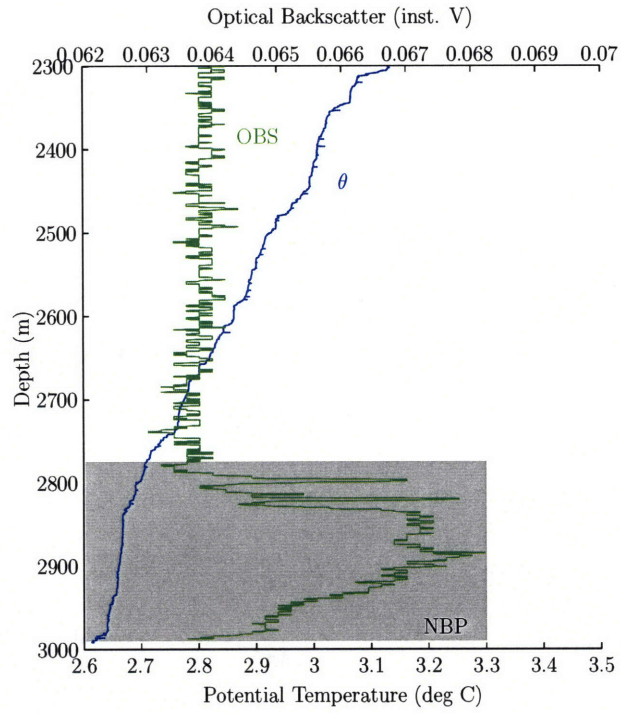


Figure F-62: Descent profiles vs. depth, ABE-151.

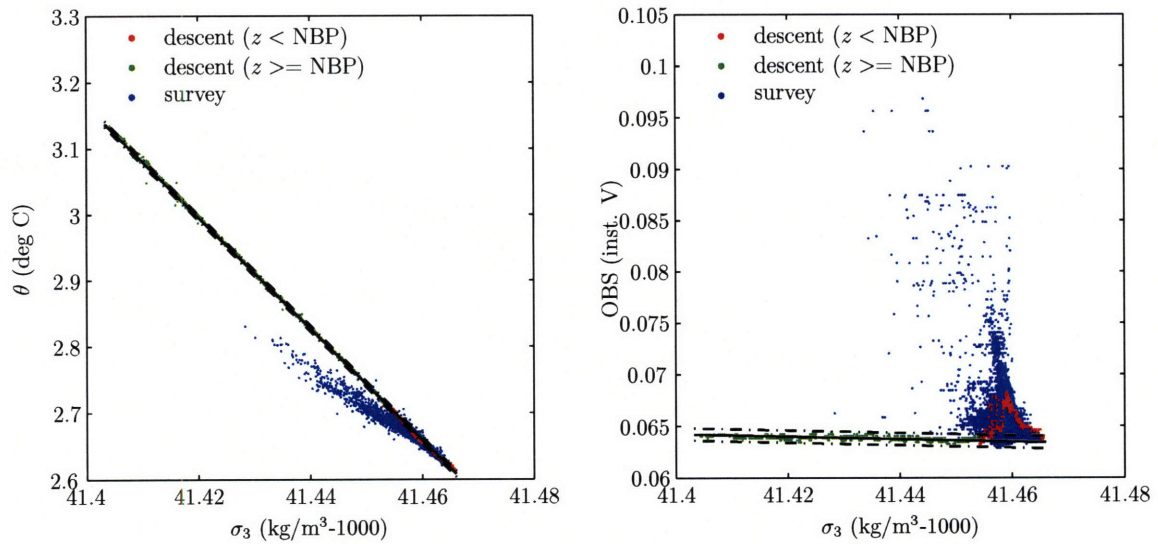


Figure F-63: Descent profiles vs. potential density, ABE-151.

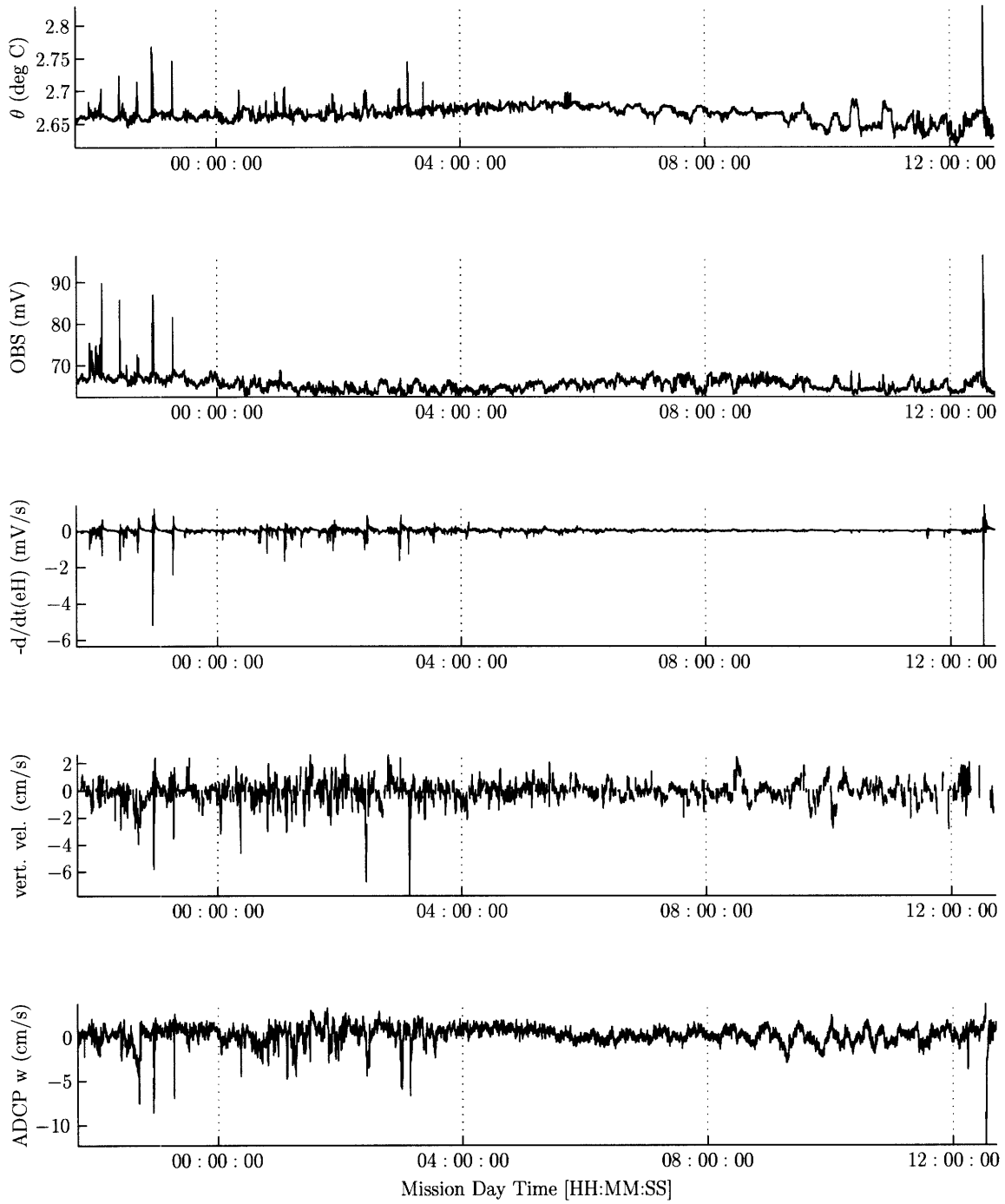


Figure F-64: Time series of hydrothermal tracers after pre-processing, ABE-151.

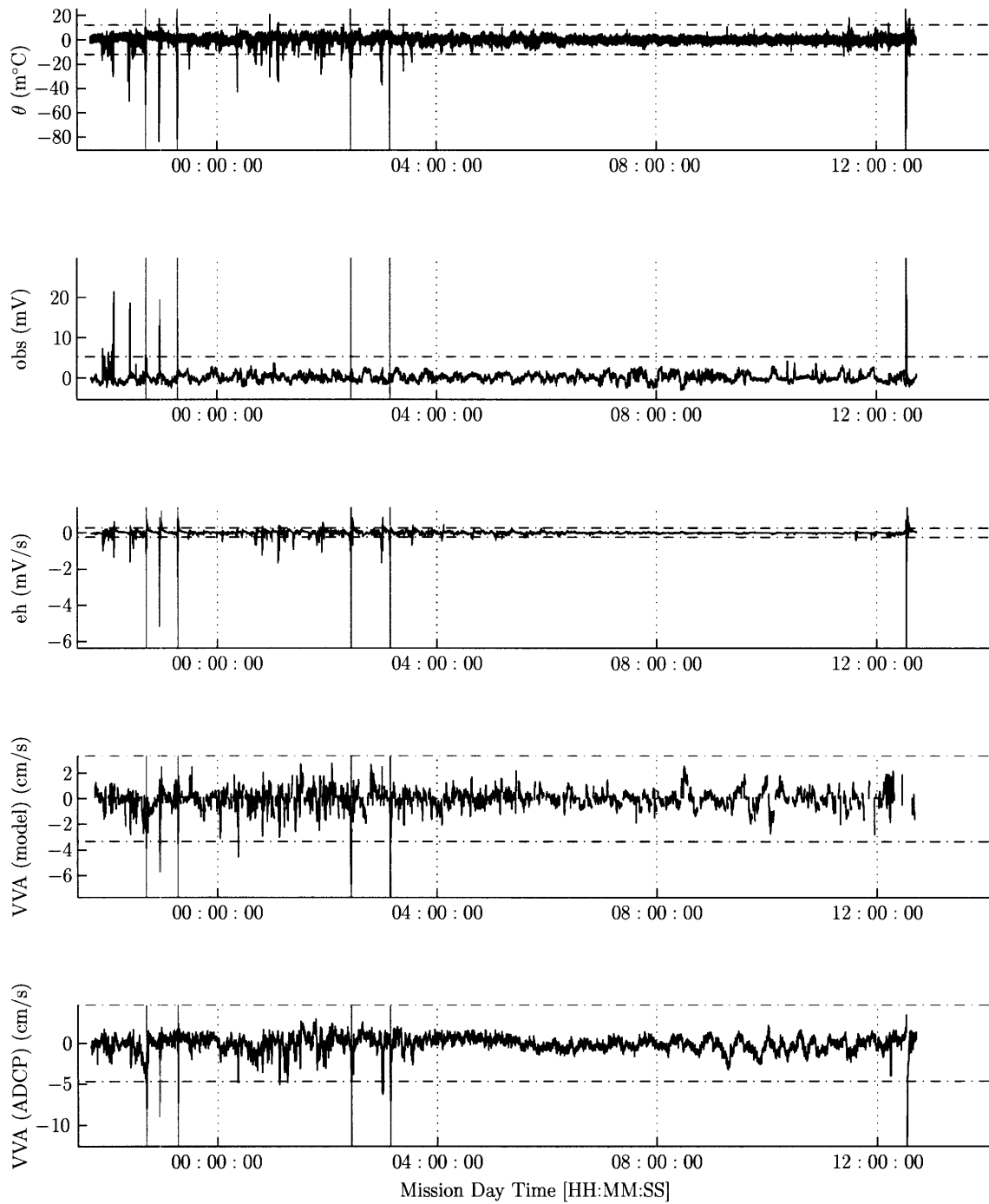


Figure F-65: Time series of hydrothermal tracers processed for BP detection, ABE-151. Intervals highlighted in gray indicate BP detection.

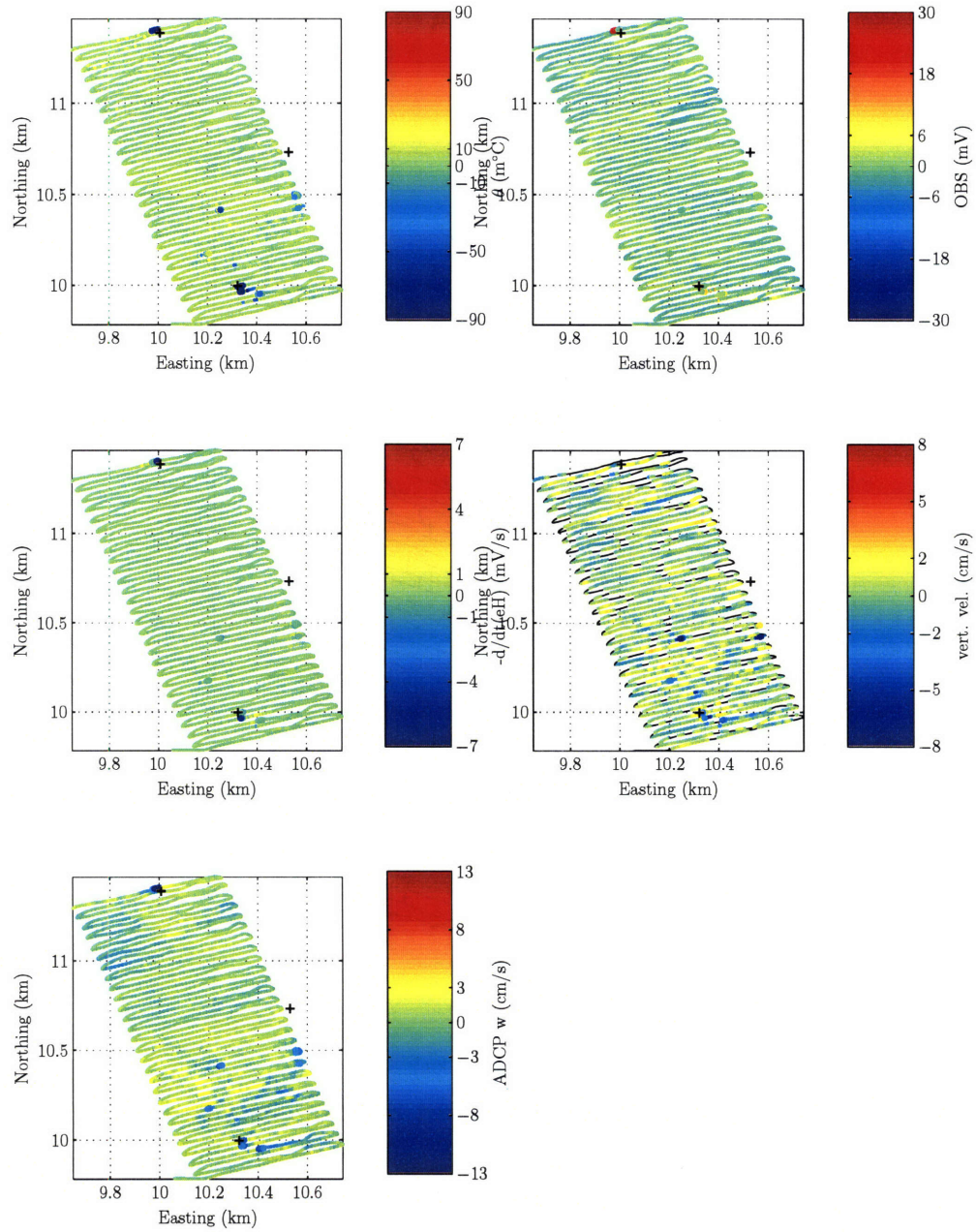


Figure F-66: Bird's eye view of anomaly intensity after processing for BP detection, ABE-151. Larger diameter dots indicate BP detection.

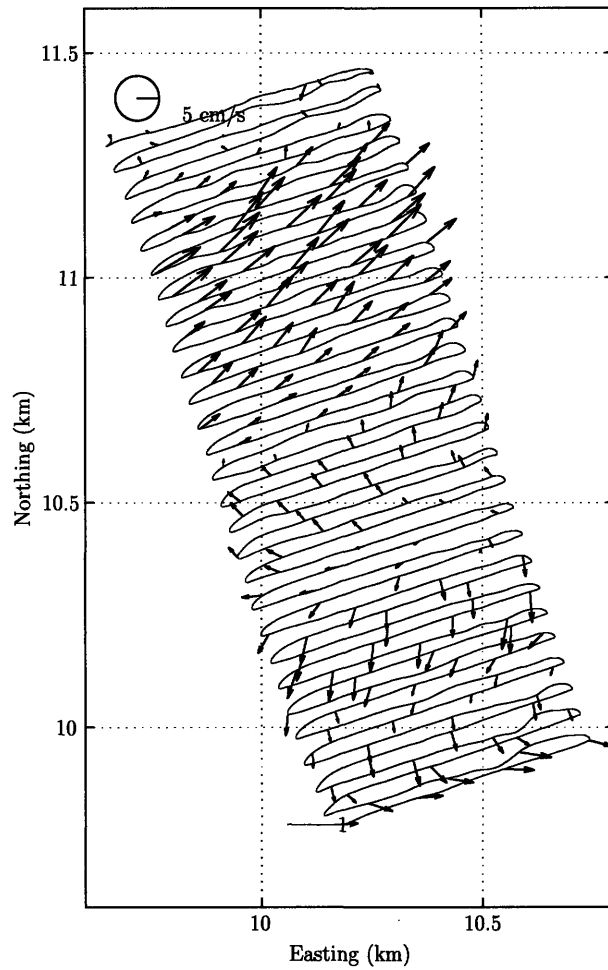


Figure F-67: Bird's eye view of ADCP-derived crossflow velocity, ABE-151 (5 min vertical average over depth interval 5 m to 25 m below vehicle). Numbers indicate the order of observation.

F.4.3 ABE-153

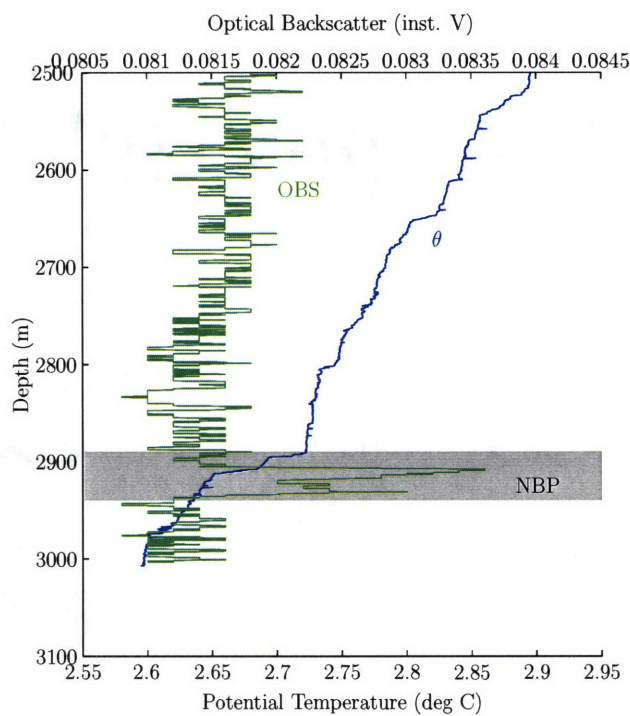


Figure F-68: Descent profiles vs. depth, ABE-153.

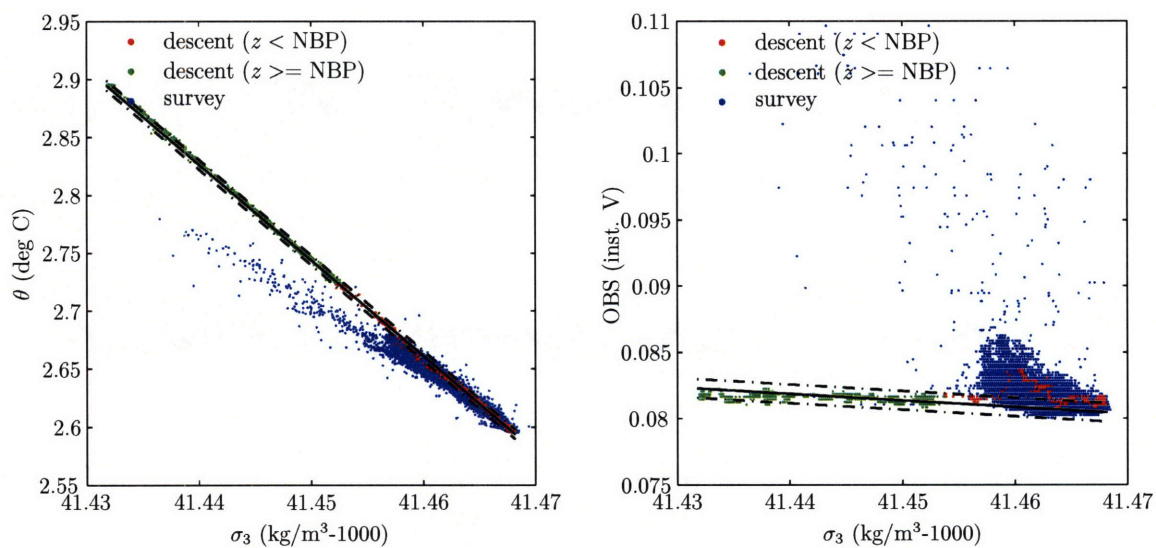


Figure F-69: Descent profiles vs. potential density, ABE-153.

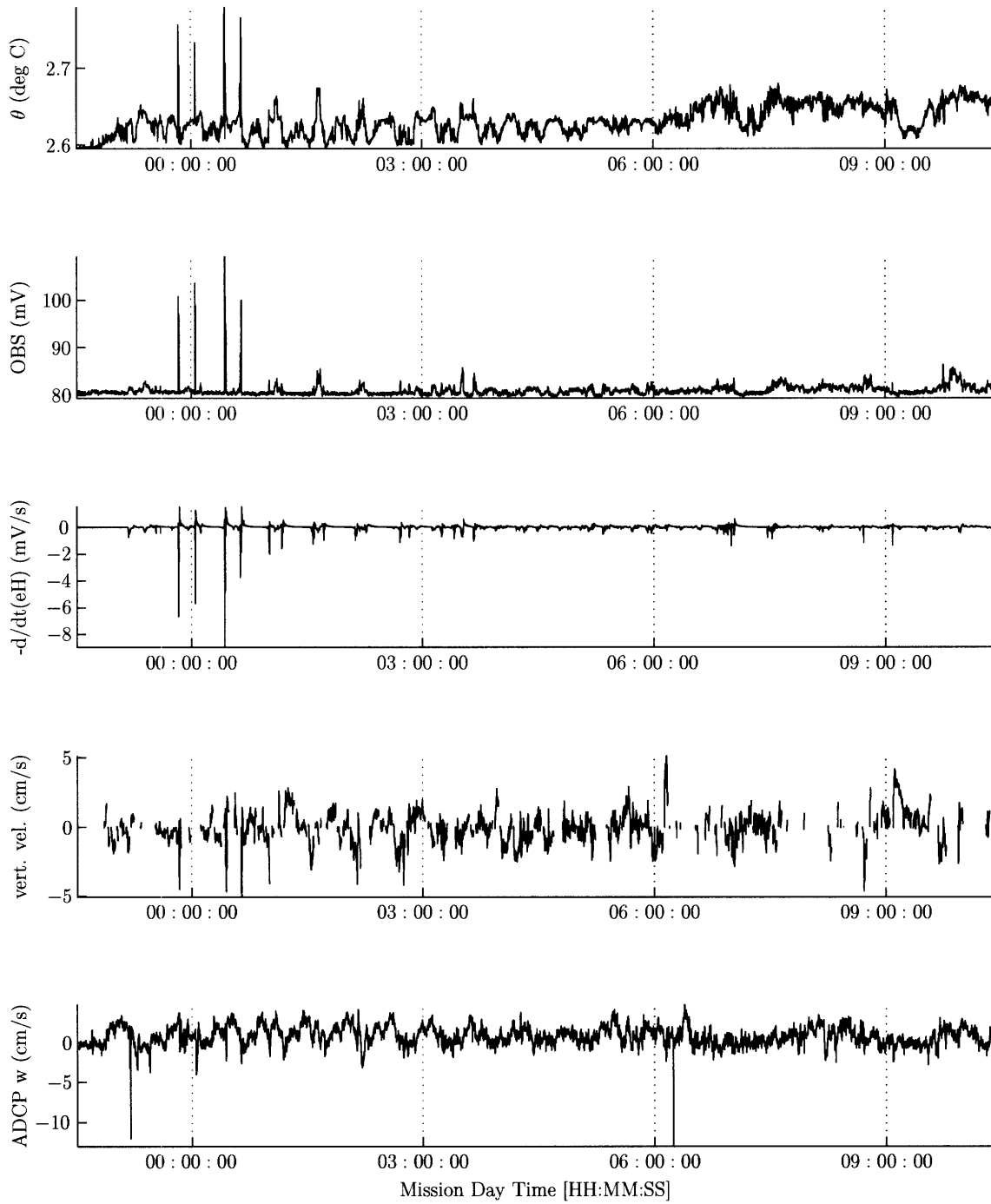


Figure F-70: Time series of hydrothermal tracers after pre-processing, ABE-153.

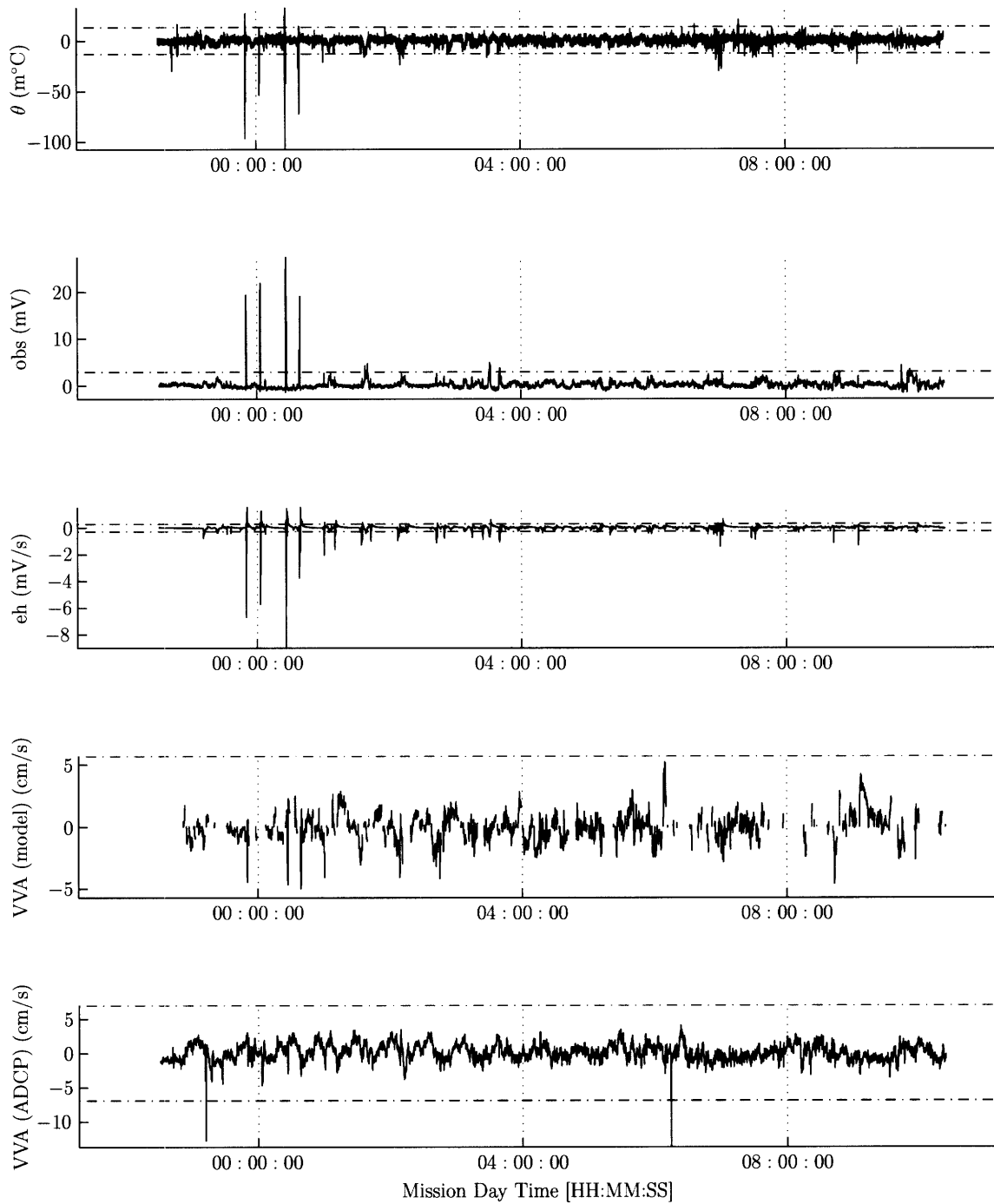


Figure F-71: Time series of hydrothermal tracers processed for BP detection, ABE-153. Intervals highlighted in gray indicate BP detection.

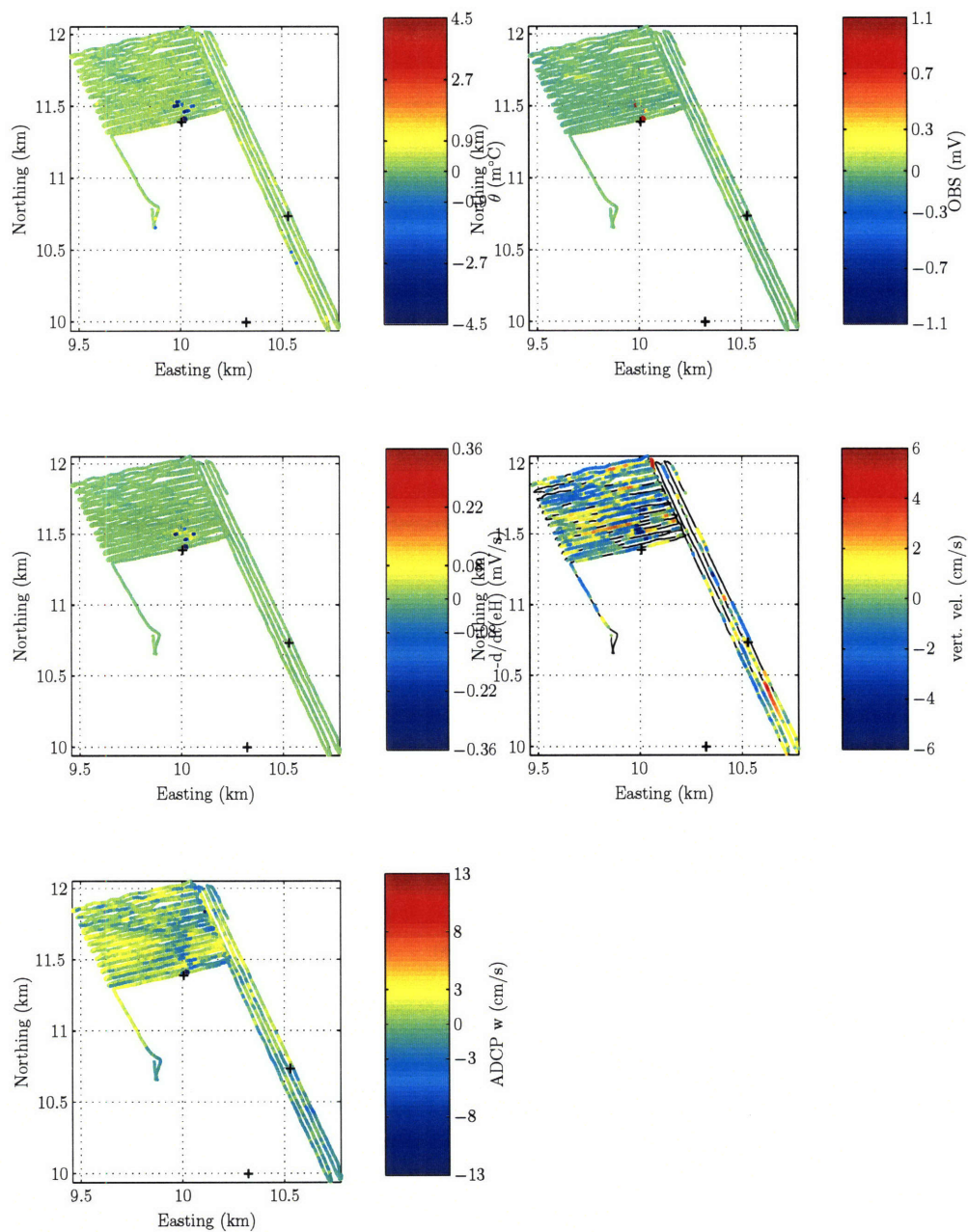


Figure F-72: Bird's eye view of anomaly intensity after processing for BP detection, ABE-153. Larger diameter dots indicate BP detection.

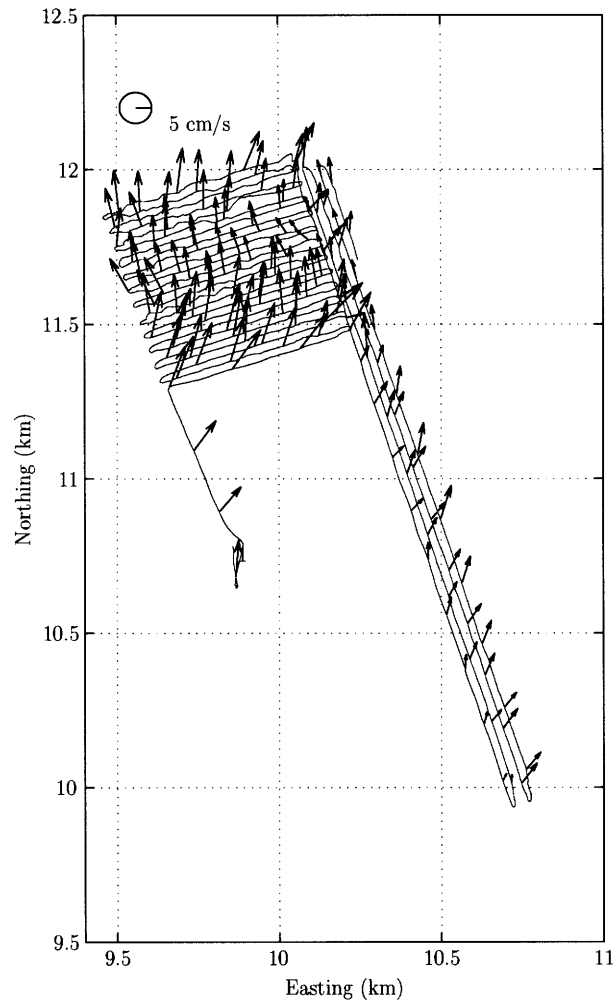


Figure F-73: Bird's eye view of ADCP-derived crossflow velocity, ABE-153 (5 min vertical average over depth interval 5 m to 25 m below vehicle). Numbers indicate the order of observation.

Appendix G

Comparison of OG Algorithms for Various Prior

The appendix explores the effect of the initializing OG maps for hydrothermal vent localization with various priors. The unique aspect of applying OG methods to the plume localization problem is the small number of cells expected to be occupied by sources. From a Bayesian perspective, this implies cells are a priori unlikely to be occupied and should therefore be initialized with small prior probabilities of occupancy. As with all Bayesian problems, the accuracy of the outcome is influenced by the accuracy of the prior. Because all practical OG methods are approximate, one might expect that the choice of OG algorithm will also have an effect on the accuracy of the result. This is indeed the case. The simulations presented in § 3.3 indicated that:

1. good results could be attained with the IP algorithm;
2. the standard algorithm is unsuitable for low prior environment.

Both these assertions are borne out here with real data from ABE-151, a Phase-2 dive (50 m height above bottom) at SMAR. In addition, since only very few BP detections were encountered during ABE-151, the exact result could also be computed,¹ pro-

¹In fact the CID algorithm was used to generate the “exact” results, with subset size and threshold set to produce an optimal grouping such that only detections spatially far apart were treated as independent.

viding the opportunity to quantitatively assess the quality of the IP approximation.

G.1 Results: ABE-151 at Nominal Prior

Figures G-1, G-2, & G-3 show OG maps constructed using the exact algorithm, the recursive form of the IP algorithm (§ 3.2.1), and the standard algorithm, respectively. The maps show the posterior odds of occupancy by the center of a vent field. These maps were initialized with a nominal prior of 10^{-5} in each $5\text{ m} \times 5\text{ m}$ grid cell corresponding to an a priori vent field density of approximately $1/\text{km}^2$.

All three maps are qualitatively similar in that they have produced peaks in the posterior odds in roughly the same places². However, the numerical values of the posterior odds in the IP map agree relatively well with the exact map in comparison to the posterior odds computed using the standard method. One practical reason to prefer the more accurate result beyond simply unease with the inaccurate posteriors produced by the standard algorithm is that the transformation from field occupancy to vent occupancy (§ 5.3.1) applied to the standard result produces huge vent fields. This happens because of the large number of cells apparently occupied by vent field centers. Furthermore, the recursive form of the IP algorithm used to produce Fig. G-2 incurs the same computational load as the standard algorithm, meaning there is no reason not to choose the IP algorithm in favor of the standard one for this application.

G.2 Algorithm Performance for Various Priors

This section explores the effect of varying the prior on the maps produced by the exact, IP, and standard OG mapping algorithms. As above the input data is from ABE-151. Figure G-4 shows that the IP algorithm produces only slight errors in cell posteriors relative to the exact result over a broad range of priors, in stark contrast to the errors produced by the standard algorithm which tend to grow as the prior is reduced.

Figure G-5 shows the expected number of source vent fields based on the posteriors produced by each algorithm. The IP and exact results agree well over the range

²Recall that all algorithms studied in this thesis behave identically with respect to isolated non-detections, thus the portions of these maps away from any detections are all identical (and exact).

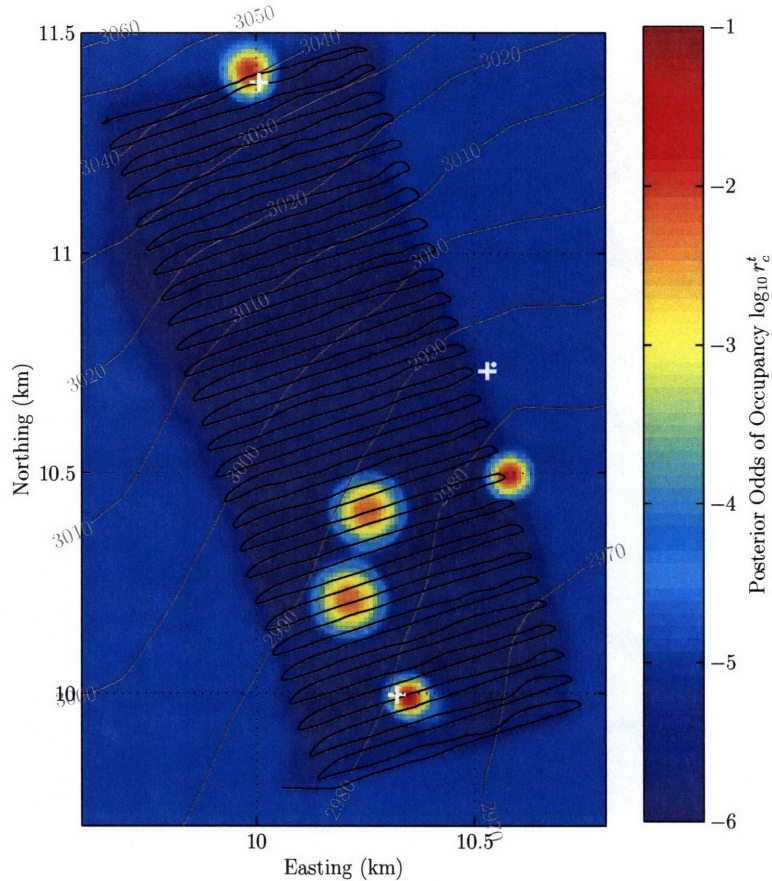


Figure G-1: Exact OG map produced from Phase-2 dive ABE-151. Only two of the five high-odds regions shown in the map were explored on subsequent dives; however, both contained active vents. The weakest and broadest peaks in the map (at coordinates (10.20,10.25) and (10.30,10.45)) correspond to single detections. Multiple detections over the remaining sites enabled the locations of these sites to be better resolved, resulting in more tightly constrained peaks with higher odds. The eastern-most vent field (starred) was discovered on a subsequent expedition over a year after ABE-151, and may not have been present during the original dive.

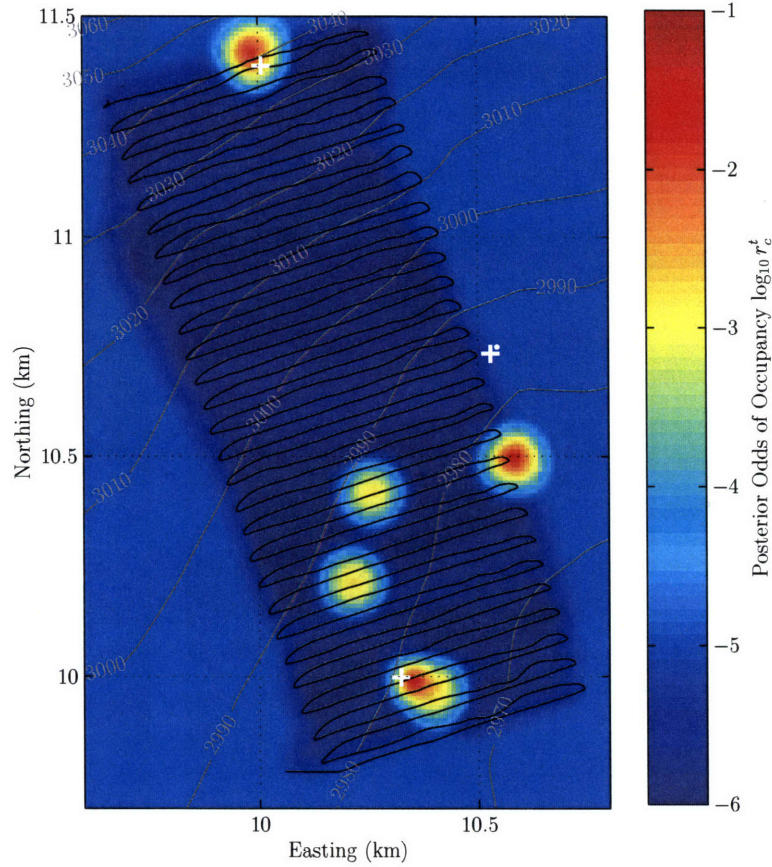


Figure G-2: IP algorithm version of the exact OG map in Fig. G-1. The map has produced odds near to the exact result, though cells with increased odds tend to spread over broader regions than in the exact map. The recursive form of the IP algorithm includes no mechanism to ever reduce the posteriors following a detection, which results in the observed broadening. Of lesser significance, cells in the two weaker peaks at coordinates (10.20,10.25) and (10.30,10.45) contain somewhat lower odds than the exact result. This is a consequence of incorporating detections into the map irreversibly as a “revised prior.” Subsequent non-detections acted reduce the posterior in these regions below that of the exact result.

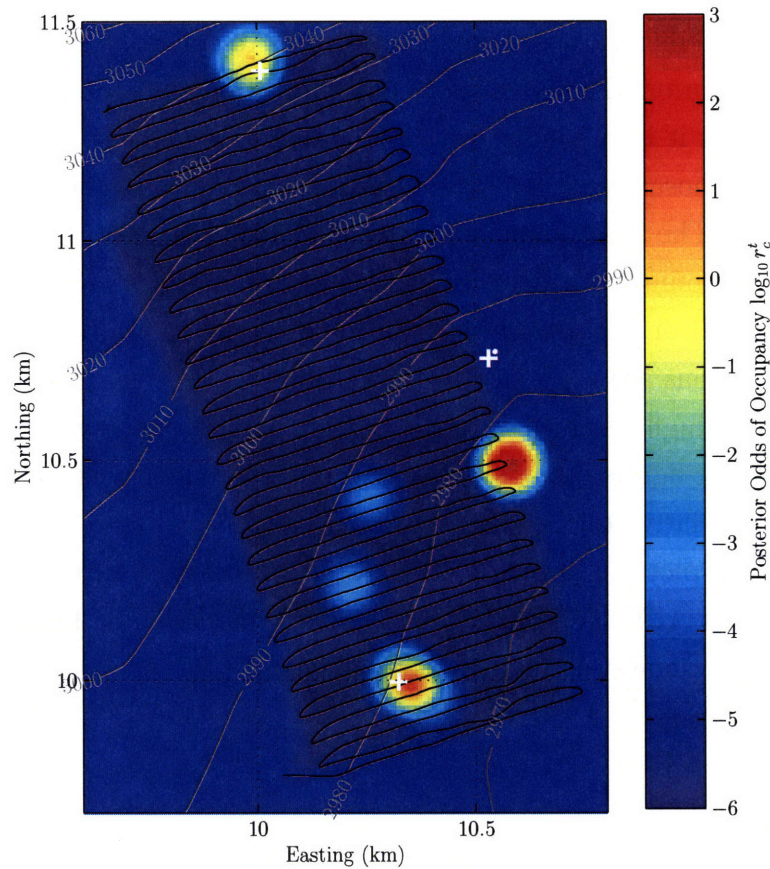


Figure G-3: Standard algorithm version of the exact OG map in Fig. G-1. The color axis has been rescaled relative to Figs. G-1 & G-2 to show the high posterior odds computed in some regions of the map. Note that these odds are indeed very high. Whereas the exact result shows maximum odds of between 0.001 and 0.01, this map shows posterior odds as high as 1000, indicating a 1000 : 1 chance that the cell is occupied. Obviously, odds this high spread over a significant number of cells are inconsistent with the low assumed prior. That said, the locations of peaks in the map agree reasonably well with the exact result.

of priors investigated, whereas the standard algorithm produces a significant overestimate for small priors. Of more interest is the behavior of the predicted number of source fields relative to the actual groundtruth number of fields. For small priors the exact and IP results predict the minimum number of source fields that must have been present in the survey area based on human interpretation of the data. As the prior increases, the number of predicted fields begins to track the number of expected fields based on the prior. That behavior is an indication that the prior was selected incorrectly, and that it is in conflict with the actual environment.

An incorrectly chosen prior has real consequences regardless of the algorithm, as shown by Fig. G-6. The figure shows the “efficiency” of each algorithm’s map in terms of the number of groundtruth field locations per cell raised above the prior by the indicated factor (1.01 or 10). Priors that are too low tend to produce inefficient maps because non-detections are weakened in their ability to constrain vent field location. Priors that are too high produce outright failure, that is, maps with no cells raised above the prior. The figure also shows the cost of the broader peaks produced by the IP approximation relative to the exact result.

Based on these plots, a prior of 10^{-5} (approximately 1 field per km^2) was a reasonable choice for ABE-151. Slightly higher performance could have been achieved with a prior of 10^{-4} , however, further increase would have produced maps near the failure point.

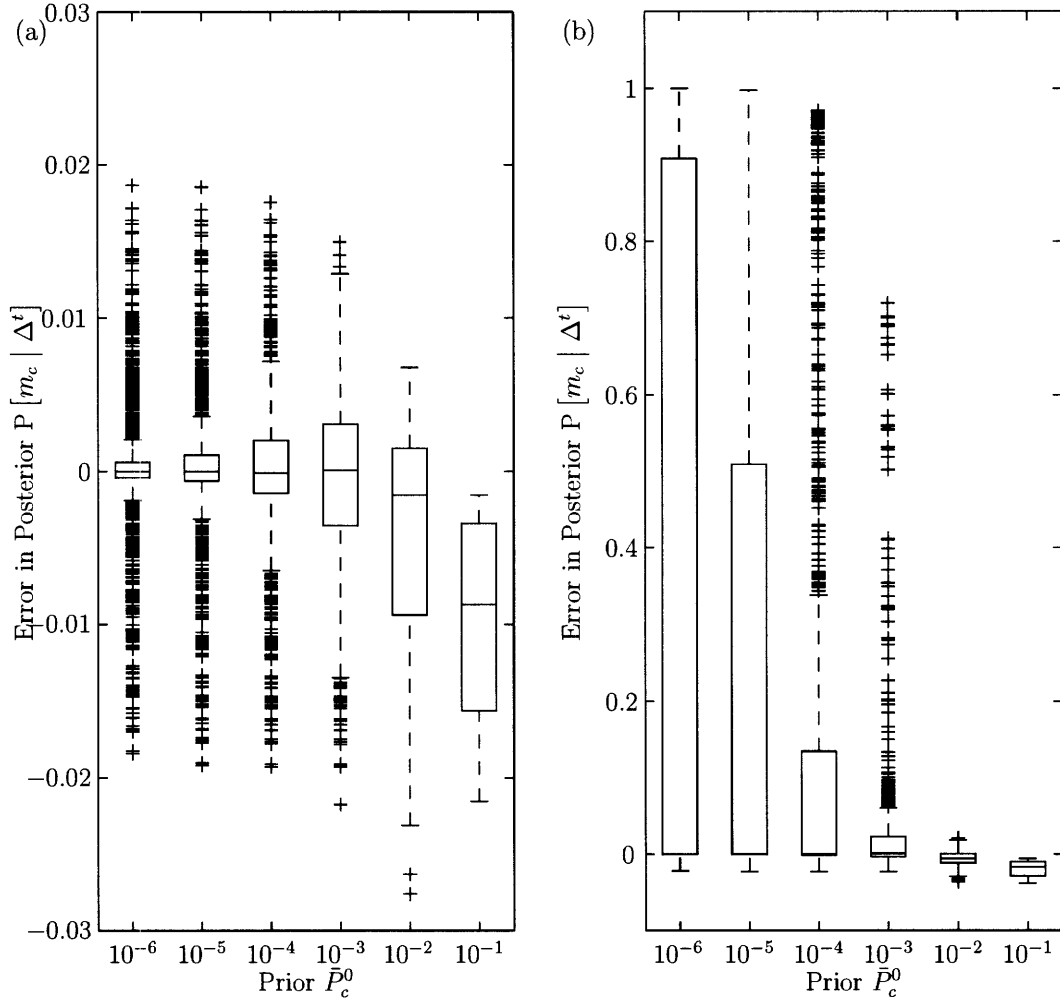


Figure G-4: Boxplots (e.g. [45]) of errors in the posteriors computed by: (a) the IP and (b) the standard OG algorithms for various priors using data from ABE-151. Note the difference in the scaling of the y-axis between the two plots. Both algorithms produce small errors relative to the exact result for relatively large priors ≥ 0.01 . For smaller priors the standard algorithm tends to produce large posteriors near unity (for the smallest priors shown the distribution of errors is actually strongly bimodal), whereas the IP algorithm exhibits fairly consistent small errors.

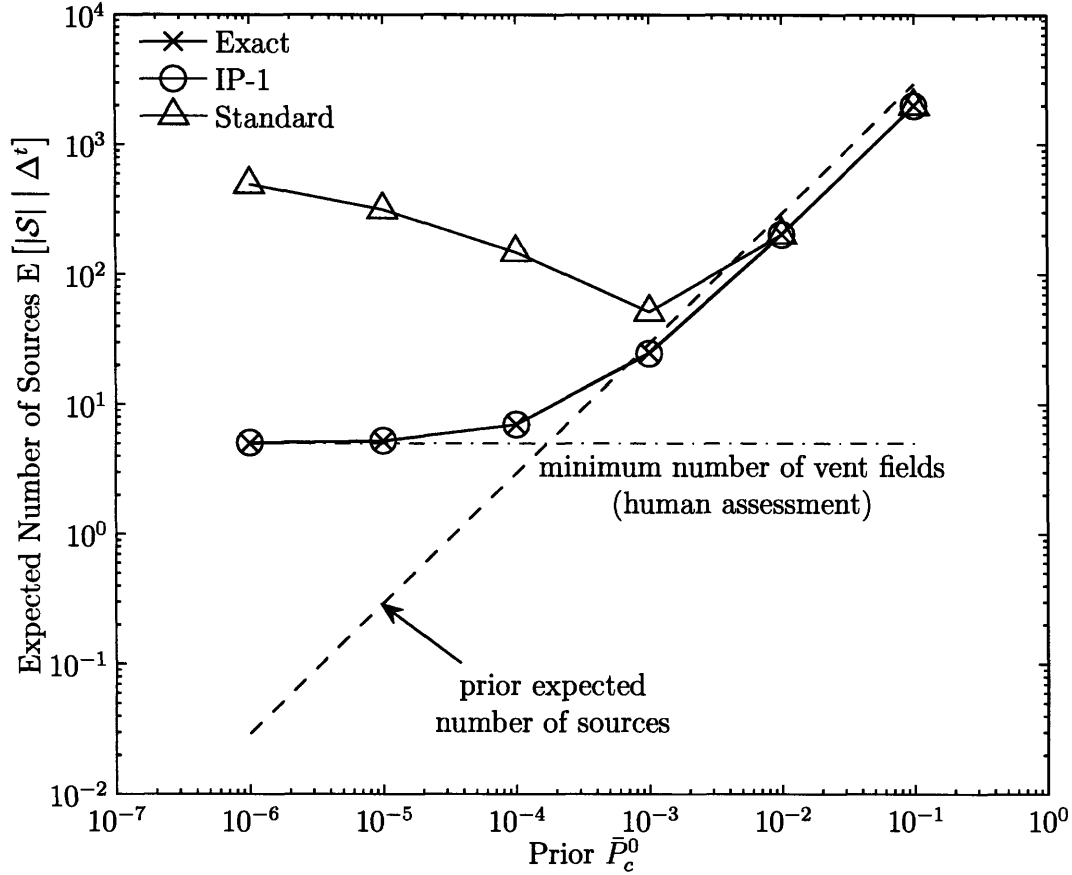


Figure G-5: Expected number of vent fields from OG maps produced by the exact, IP and standard algorithms for various priors. The IP and exact results agree well over the range of priors investigated, whereas the standard algorithm produces a significant overestimate for small priors. Interestingly, as the prior increases beyond the minimum number of fields inferred by human interpretation of the data, the exact and IP results tend to track the prior predicted number of source fields. This behavior illustrates the susceptibility of all Bayesian methods to errors induced by an incorrect prior. It might be inferred from this plot that priors well below the actual will not compromise the accuracy of the exact and IP results, whereas high priors might. To some extent this is true; however, the utility of non-detections will be decreased artificially because detections become even less likely. As a result, peaks in the posterior will be broader and the map less “efficient,” as is indicated by Fig. G-6. In addition, the dashed horizontal line in the plot indicates only the minimum number of fields. Recall that these fields were assumed to have a characteristic number of sources Q each with characteristic size b_0 . In this particular case (ABE-151), detections were more numerous than expected (cf. Table 5.1), so that what appears as an over-estimate in the plot around a prior of 10^{-4} instead indicates vent fields with more individual vents or a larger characteristic source area than expected.

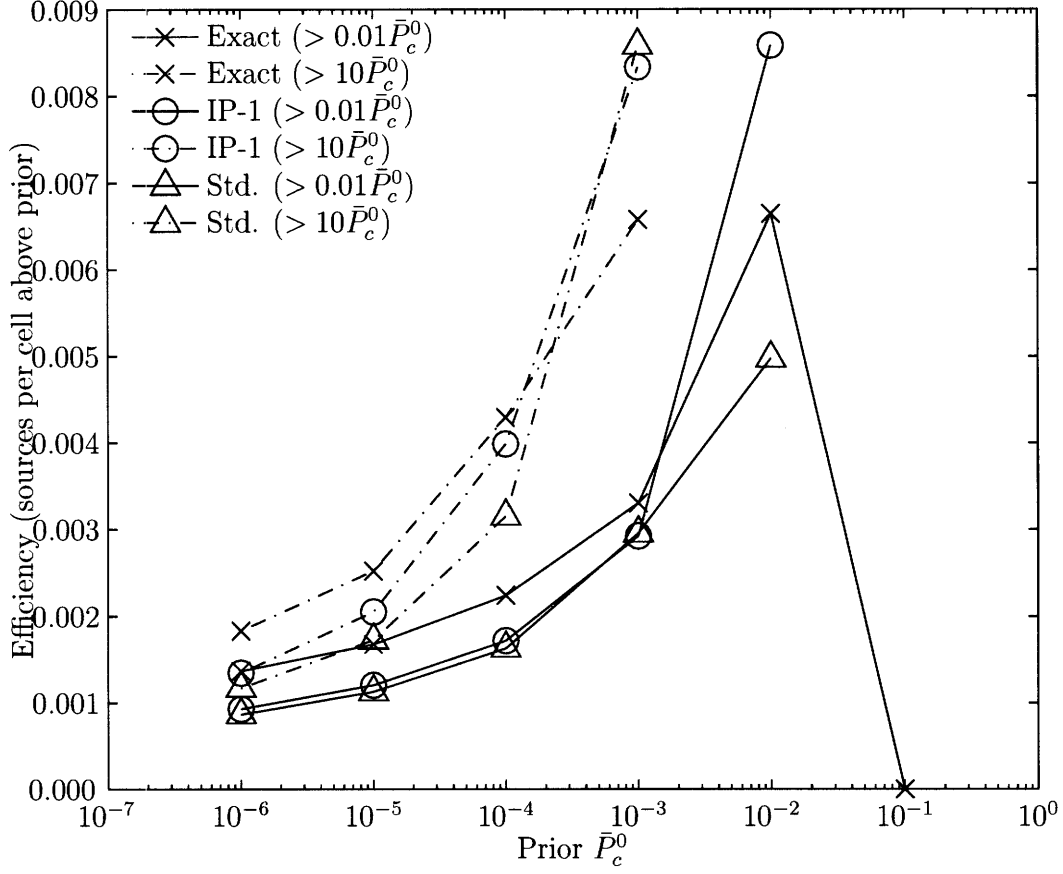


Figure G-6: OG algorithm efficiency for various priors. Efficiency is defined as the number of groundtruth field locations per cell raised above the prior by the indicated factor. The small factor (1.01) includes essentially all cells that were raised above the prior at all, while the large factor (10) includes only the cells most likely to contain the centers of source fields. Higher efficiencies were attained with the large factor indicating good agreement with groundtruth source locations. Efficiency suffers for priors that are too low, whereas outright failure (no cells above the prior) occurs for priors that are too high. The exact algorithm tends to produce the best maps, except near the point where outright failure occurs.

Bibliography

- [1] Ben Allen, Roger Stokey, Tom Austin, Ned Forrester, Rob Goldsborough, Mike Purcell, and Chris von Alt. REMUS: A small, low cost AUV; system description, field trials and performance results. In *Proceedings IEEE OCEANS 1997*, pages 994–1000, 1997.
- [2] J. Atema. Chemical signals in the marine environment: dispersal, detection, and temporal signal analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 92:62–66, 1995.
- [3] J. Atema. Eddy chemotaxis and odor landscapes: exploration of nature with animal sensors. *Biological Bulletin*, 191(1):129–138, 1996.
- [4] E. T. Baker. *Gorda Ridge: A Seafloor Spreading Center in the United States Exclusive Economic Zone*, chapter Hydrothermal plume prospecting: hydrographic and geochemical techniques, pages 155–167. Springer-Verlag, New York, 1990.
- [5] Edward T. Baker and Christopher R. German. *Mid-Ocean Ridges: Hydrothermal Interactions Between the Lithosphere and Oceans*, chapter On the Global Distribution of Hydrothermal Vent Fields, pages 245–266. Geophysical Monograph 148. American Geophysical Union, 2004.
- [6] Edward T. Baker, Christopher R. German, and Henry Elderfield. *Seafloor Hydrothermal Systems: Physical, Chemical, Biological, and Geological Interactions*, chapter Hydrothermal Plumes Over Spreading-Center Axes: Global Distributions and Geological Inferences, pages 47–71. Geophysical Monograph 91. American Geophysical Union, 1995.
- [7] E.T. Baker and H. B. Milburn. MAPR: A new instrument for hydrothermal plume mapping. *Ridge Events*, 8(1):23–25, January 1997.
- [8] Eugene Balkovsky and Boris I. Shraiman. Olfactory search at high Reynolds number. In *Proceedings National Academy of Science*, volume 99, pages 12589–12593, October 2002.

- [9] J. H. Belanger and M. A. Willis. Adaptive control of odor guided locomotion: behavioral flexibility as an antidote to environmental unpredictability. *Adaptive Behavior*, 4:217–253, 1996.
- [10] Ami Berler and Solomon E. Shimony. Bayes networks for sensor fusion in occupancy grids. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 1997.
- [11] Erik Burian, Dana Yoerger, Albert Bradley, and Hanumant Singh. Gradient search with autonomous underwater vehicles using scalar measurements. *Proc. AUV'96*, pages 86–98, 1996.
- [12] R. Camilli, B. Bingham, M. Jakuba, H. Singh, and J. Whelan. Integrating in-situ chemical sampling with AUV control systems. In *OCEANS 2004, MTS/IEEE Oceans Conference*, 2004.
- [13] R. Camilli and H. F. Hemond. NEREUS/Kemonaut, a mobile autonomous underwater mass spectrometer. *Trends in Analytical Chemistry*, 23(4):307–313, 2004.
- [14] Michael J. Caruso. Applications of magnetic sensors for low cost compass systems. In *Position Location and Navigation Symposium, IEEE 2000*, pages 177–184, March 2000.
- [15] Vassilios N. Christopoulos and Stergios Roumeliotis. Adaptive sensing for instantaneous gas release parameter estimation. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 4450–4456, Barcelona, Spain, April 2005.
- [16] Vassilios N. Christopoulos and Stergios Roumeliotis. Multi robot trajectory generation for single source explosion parameter estimation. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 2803–2809, Barcelona, Spain, April 2005.
- [17] D. P. Connelly, R. D. Prien, C. R. German, and P. J. Statham. Development of a multi-platform in situ analyser for dissolved iron and manganese. In *EGU General Assembly*, Vienna, Austria, 2005. Abstract EGU05-A-01655.
- [18] T. R. Consi, J. Atema, C. A. Goudey, J. Cho, and C. Chrysostomidis. Auv guidance with chemical signals. In *Proceedings of the 1994 Symposium on Autonomous Underwater Vehicle Technology (AUV'94)*, pages 450–455, July 1994.

- [19] J.B. Corliss, J. Dymond, Louis I. Gordon, John M. Edmond, Richard P. von Herzen, Robert D. Ballard, Kenneth Green, David Williams, Arnold Bainbridge, Kathy Crane, and Tjeerd H. van Andel. Submarine thermal springs on the Galapagos Rift. *Science*, 203:1073–1083, 1979.
- [20] G. T. Csanady. *Turbulent Diffusion in the Environment*. D. Reidel Publishing Co., Dordrecht, The Netherlands, 1973.
- [21] Jonathan Dale. Chemosensory search behavior in the starfish *asterias forbesi*. *Biological Bulletin*, 193(2):210–212, 1997.
- [22] Laurie Davies and Ursula Gather. The identification of multiple outliers. *Journal of the American Statistical Association*, 88:782–792, September 1993.
- [23] T. Duckett, M. Axelsson, and A. Saffiotti. Learning to locate an odour source with a mobile robot. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'2001)*, Seoul, Korea, 2001.
- [24] D. B. Dusenbery. Performance of basic strategies for following gradients in two dimensions. *Journal of Theoretical Biology*, 208:345–360, 2001.
- [25] Alberto Elfes. Sonar-based real-world mapping and navigation. *IEEE Transactions on Robotics and Automation*, 3(3):249–265, June 1987.
- [26] Alberto Elfes. *Occupancy Grids: A Probabilistic Framework for Mobile Robot Perception and Navigation*. PhD thesis, Carnegie-Mellon University, Electrical and Computer Engineering Dept., 1989.
- [27] Alberto Elfes. Using Occupancy Grids for mobile robot perception and navigation. *Computer*, 22(6):46–57, June 1989.
- [28] Ahmed Mohamod Farah and Tom Duckett. Reactive localisation of an odour source by a learning mobile robot. In *Proceedings of the Second Swedish Workshop on Autonomous Robotics*, Stockholm, Sweden, October 10-11, 2002.
- [29] J. Farrell, W. Li, S. Pang, and R. Arrieta. Chemical plume tracing experimental results with a REMUS AUV. In *Oceans 2003 Marine Technology and Ocean Science Conference*, San Diego, CA, September 22-26 2003.
- [30] J. A. Farrell, J. Murlis, X. Long, W. Li, and R. T. Cardé. Filament-based atmospheric dispersion model to achieve short time-scale structure of odor plumes. *Environmental Fluid Mechanics*, 12:143–169, 2002.
- [31] J. A. Farrell, S. Pang, and W. Li. Chemical plume tracing via an autonomous underwater vehicle. *IEEE Journal of Oceanic Engineering*, 2005. To appear.

- [32] Jay A. Farrell, Shuo Pang, and Wei Li. Plume mapping via hidden markov methods. *IEEE Transactions on System, Man, and Cybernetics*, 33(6):850–863, 2003.
- [33] Gabriele Ferri, Emanuele Caselli, Virgilio Mattoli, Alessio Mondini, Barbara Mazzolai, and Paolo Dario. A biologically-inspired algorithm implemented on a new highly flexible multi-agent platform for gas source localization. In *Proceedings of The First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics, BioRob 2006*, pages 573–578, February 2006.
- [34] C. M. Finelli, N. D. Pentcheff, R. Z. Zimmer-Faust, and D. S. Wetthey. Odor transport in turbulent flows: constraints on animal navigation. *Limnology and Oceanography*, 44:1056–1071, 1999.
- [35] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, June 1981.
- [36] N. P. Fofonoff and R. C. Millard Jr. Algorithms for computation of fundamental properties of seawater. Technical Report 44, Unesco technical papers in marine science, 1983. pp. 53.
- [37] C. German. Personal Communication, September 2006.
- [38] C. German, D. Connelly, R. Prien, D. Yoerger, M. Jakuba, A. Bradley, T. Shank, K. Nakamura, H. Edmonds, and C. Langmuir. New techniques for hydrothermal exploration: In situ chemical sensors on AUVs preliminary results from the Lau Basin. In *Eos. Trans. AGU*, volume 85, 2004. Fall Meet. Suppl., Abstract B13A-0190.
- [39] C. R. German, D. P. Connelly, R. D. Prien, D. R. Yoerger, M. Jakuba, A. Bradley, T. Shank, K. Nakamura, C. Langmuir, and L. M. Parsons. New techniques for hydrothermal plume investigation by AUV. In *Geophysical Research Abstracts*, volume 7, Vienna, Austria, 2005. European Geosciences Union.
- [40] Christopher R. German, Jian Lin, and Lindsay M. Parson, editors. *Mid-Ocean Ridges: Hydrothermal Interactions Between the Lithosphere and Oceans*. Geophysical Monograph 148. American Geophysical Union, 2004.
- [41] Frank W. Grasso. Invertebrate-inspired sensory-motor systems and autonomous, olfactory-guided exploration. *Biological Bulletin*, 200:160–168, 2001.
- [42] Frank W. Grasso and Jelle Atema. Integration of flow and chemical sensing for guidance of autonomous marine robots in turbulent flows. *Environmental Fluid Mechanics*, 2:95–114, 2002.

- [43] Frank W. Grasso, Thomas R. Consi, David C. Mountain, and Jelle Atema. Biomimetic robot lobster performs chemo-orientation in turbulence using a pair of spatially separated sensors: progress and challenges. *Robotics and Autonomous Systems*, 30:115–131, 2000.
- [44] Adam T. Hayes, Alcherio Martinoli, and Rodney M. Goodman. Distributed Odor Source Localization. *IEEE Sensors*, June 2002.
- [45] Anthony J. Hayter. *Probability and Statistics for Engineers and Scientists*. Duxbury Press, 2 edition, 2002.
- [46] Karl R. Helfrich and Kevin G. Speer. *Seafloor Hydrothermal Systems: Physical, Chemical, Biological, and Geological Interactions*, chapter Oceanic Hydrothermal Circulation: Mesoscale and Basin-Scale Flow, pages 347–356. Geophysical Monograph 91. American Geophysical Union, 1995.
- [47] B. Henderson-Sellers. Shape constants for plume models. *Boundary-Layer Meteorology*, 21:105–114, 1981.
- [48] A. Howard and L. Kitchen. Generating sonar maps in highly specular environments, 1996.
- [49] Susan E. Humphris, Robert A. Zierenberg, Lauren S. Mullineaux, and Richard E. Thomson, editors. *Seafloor Hydrothermal Systems: Physical, Chemical, Biological, and Geological Interactions*. Geophysical Monograph 91. American Geophysical Union, 1995.
- [50] J. Ishibashi, J. E. Lupton, T. Yamaguchi, J. Querellou, T. Nunoura, and K. Takai. Expedition reveals changes in Lau Basin hydrothermal system. *EOS Transactions, AGU*, 87(2):13–24, January 2006.
- [51] H. Ishida, T. Nakamoto, and T. Moriizumi. Remote sensing of gas/odor source location and concentration distribution using mobile system. *Sensors and Actuators B*, 49:52–57, 1998.
- [52] H. Ishida, T. Nakamoto, T. Moriizumi, T. Kikas, and J. Janata. Plume-tracking robots: a new application of chemical sensors. *Biological Bulletin*, 200(2):222–226, April 2001.
- [53] D. C. Kadko, N. D. Rosenberg, J. E. Lupton, R. W. Collier, and M. D. Lilley. Chemical reaction rates and entrainment within the Endeavour Ridge hydrothermal plume. *Earth and Planetary Science Letters*, 99:315–335, 1990.
- [54] K.E. Kaissling. *Orientation and Communication in Arthropods*, chapter Pheromone-controlled anemotaxis in moths. Birkhäuser, Basel, 1997.

- [55] Deborah S. Kelley, Jeffrey A. Karson, Donna K. Blackman, Gretchen L. Fröh-Green, David A. Butterfield, Marvin D. Lilley, Eric J. Olson, Matthew O. Schrenk, Kevin K. Roe, Geoff T. Lebon, Pete Rivizzigno, and & the AT3-60 Shipboard Party. An off-axis hydrothermal vent field near the Mid-Atlantic Ridge at 30 deg N. *Nature*, 412, July 2001.
- [56] James C. Kinsey and Louis L. Whitcomb. In-situ alignment calibration of attitude and Doppler sensors for precision underwater vehicle navigation: Theory and experiment. *IEEEJOE*, 2006. In print.
- [57] John A. Knauss. *Introduction to Physical Oceanography*. Prentice Hall, 2nd edition, 1996.
- [58] Kurt Konolige. Improved occupancy grids for map building. *Autonomous Robots*, 4:351–367, 1997.
- [59] Gerhard K. Kraetzschmar, Guillem Pagès Gassull, and Klaus Uhl. Probabilistic quadrees for variable-resolution mapping of large environments. In M. I. Ribeiro and J. Santos Victor, editors, *Proceedings of the 5th IFAC/EURON Symposium on Intelligent Autonomous Vehicles*, Lisbon, Portugal, July 2004. Elsevier Science.
- [60] Y. Kuwana, S. Nagasawa, I. Shimoyama, and R. Kanzaki. Synthesis of the pheromone-oriented behaviour of silkworm moths by a mobile robot with moth antennae as pheromone sensors. *Biosensors and Bioelectronics*, 14:195–202, 1999.
- [61] Y. Kuwana and I. Shimoyama. A pheromone-guided mobile robot that behaves like a silkworm moth with living antennae as pheromone sensors. *International Journal of Robotics Research*, 17(9):924–933, 1998.
- [62] Yoshihiko Kuwana, Isao Shimoyama, and Hirofumi Miura. Steering control of a mobile robot using insect antennae. In *Proceedings of the 1995 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 530–535, 1995.
- [63] Yoshihiko Kuwana, Isao Shimoyama, Yushi Sayama, and Hirofumi Miura. Synthesis of pheromone-oriented emergent behavior of a silkworm moth. In *Proceedings of the 1996 IEEE/RSJ International Conference on Intelligent Robots and Systems IROS'96*, volume 3, pages 1722–1729, 1996.
- [64] C. H. Langmuir, H. N. Edmonds, P. J. Michael, D. R. Yoerger, A. M. Bradley, T. M. Shank, S. L. Goldstein, and D. W. Graham. Integrated hydrothermal and petrological studies of the Eastern Lau Spreading Center.

- [65] C. H. Langmuir, C. German, P. Michael, D. R. Yoerger, D. J. Fornari, T. M. Shank, P. D. Asimow, H. N. Edmonds, and Lau2TEAM. Hydrothermal prospecting and petrological sampling in the Lau Basin: Background data for the integrated study site. In *Eos. Trans. AGU*, volume 85, 2004. Fall Meet. Suppl., Abstract B13A-0189.
- [66] J. W. Lavelle. Buoyancy-driven plumes in rotating, stratified cross-flows: plume dependence on rotation, turbulent mixing, and cross-flow strength. *Journal of Geophysical Research*, 102:3405–3420, 1997.
- [67] J. W. Lavelle, M. A. Wetzler, E. T. Baker, and R. W. Embley. Prospecting for hydrothermal vents using moored current and temperature data: Axial Volcano on the Juan de Fuca Ridge, Northeast Pacific. *Journal of Physical Oceanography*, 31, March 2001.
- [68] J. J. Leonard and H. F. Durrant-Whyte. *Directed sonar sensing for mobile robot navigation*. Kluwer, Dordrecht, The Netherlands, 1992.
- [69] Wei Li, J. A. Farrell, and R. T. Cardé. Tracking of fluid-advected odor plumes: Strategies inspired by insect orientation to pheromone. *Adaptive Behavior*, 9(3/4):143–170, 2001.
- [70] A. J. Lilienthal and T. Duckett. Experimental analysis of smelling braintenberg vehicles. In *Proceedings of the IEEE International Conference on Advanced Robotics (ICAR 2003)*, Coimbra, Portugal, 2003.
- [71] A. J. Lilienthal, M. R. Wandel, U. Weimar, and A. Zell. Sensing odour sources in indoor environments without a constant airflow by a mobile robot. In *International Conference on Robotics & Automation*, pages 4005–4010, 2001.
- [72] Achim Lilienthal and Andreas Zell. Experiences using gas sensors on an autonomous mobile robot. In *Proceedings of EUROBOT 2001, 4th European Workshop on Advanced Mobile Robots*, pages 1–8, 2001.
- [73] Marvin D. Lilley, Richard A. Feely, and John H. Trefry. *Seafloor Hydrothermal Systems: Physical, Chemical, Biological, and Geological Interactions*, chapter Chemical and Biochemical Transformations in Hydrothermal Plumes, pages 357–368. Geophysical Monograph 91. American Geophysical Union, 1995.
- [74] Jong Hwan Lim and Dong Woo Cho. Physically based sensor modeling for a sonar map in a specular environment. In *Proc. IEEE International Conference on Robotics and Automation*, volume 2, pages 1714–1719, May 1992.
- [75] Tony Ludlow. Modeling mechanisms of animal behavior, January 2004.

- [76] R. G. Lueck. Thermal inertia of conductivity cells: Theory. *Journal of Atmospheric and Oceanic Technology*, 7:741–755, 1990.
- [77] J. E. Lupton, J. R. Delaney, H. P. Johnson, and M. K. Tivey. Entrainment and vertical transport of deep-ocean water by buoyant hydrothermal plumes. *Nature*, 316(15):621–623, August 1985.
- [78] John E. Lupton. *Seafloor Hydrothermal Systems: Physical, Chemical, Biological, and Geological Interactions*, chapter Hydrothermal Plumes: Near and Far Field, pages 317–346. Geophysical Monograph 91. American Geophysical Union, 1995.
- [79] Martin C. Martin and Hans Moravec. Robot evidence grids. Technical Report CMU-RI-TR-96-06, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, March 1996.
- [80] F. Martinez, B. Taylor, J. A. Resing, E. Baker, S. L. Walker, and G. J. Massoth. Geophysical and hydrothermal survey of the Lau Basin Integrated Studies Site. In *Eos. Trans. AGU*, volume 85, 2004. Fall Meet. Suppl., Abstract B13A-0188.
- [81] T. J. McDougall. Bulk properties of hot smoker plumes. *Earth and Planetary Science Letters*, 99:185–194, 1990.
- [82] Russell E. McDuff. *Seafloor Hydrothermal Systems: Physical, Chemical, Biological, and Geological Interactions*, chapter Physical Dynamics of Deep-Sea Hydrothermal Plumes, pages 357–368. Geophysical Monograph 91. American Geophysical Union, 1995.
- [83] Russell E. McDuff, Dana R. Yoerger, and Albert M. Bradley. Precise Measurement of the Heat Flux from a Hydrothermal Vent System.
- [84] J. H. Middleton and R. E. Thomson. Modeling the rise of hydrothermal plumes. Technical Report 69, Canadian Technical Report of Hydrography and Ocean Sciences, 1986.
- [85] Hans P. Moravec. Sensor fusion in certainty grids for mobile robots. *AI Magazine*, 9:61–74, 1988.
- [86] Hans P. Moravec and Dong Woo Cho. A bayesian method for certainty grids. In *Working notes of AAAI 1998 Spring Symposium on Robot Navigation*, pages 57–60, Stanford, CA, 1989.
- [87] Hans P. Moravec and Alberto Elfes. High resolution maps from wide-angle sonar. In *Proc. IEEE International Conference on Robotics and Automation*, 1985.

- [88] Phillip P. Morgan. SEAWATER: A library of MATLAB computational routines for the properties of sea water. Technical Report Report 222, CSIRO Marine Laboratories, 1994. 29pp.
- [89] B. R. Morton, G. I. Taylor, and J. S. Turner. Turbulent gravitational convection from maintained and instantaneous sources. *Proc. R. Soc. London A*, 234:1–23, 1956.
- [90] J. Murlis, J.S. Elkington, and R.T. Cardé. Odor plumes and how insects use them. *Annual Review of Entomology*, 37:505–532, 1992.
- [91] J. Murlis, M. A. Willis, and R. T. Cardé. Spatial and temporal structure of pheromone plumes in fields and forests. *Physiological Entomology*, 25:211–222, 2000.
- [92] Sumito Nagasawa, Ryohei Kanzaki, and Isao Shimoyama. Study of a small mobile robot that uses living insect antennae as pheromone sensors. In *Proceedings of the 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 555–560, 1999.
- [93] Sumito Nagasawa, Ryohei Kanzaki, and Isao Shimoyama. An evaluation method for insect sensorimotor models using an insect-size mobile robot. In *Proceedings of the 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems IROS'01*, pages 551–556, 2001.
- [94] K. Nakamura. Personal Communication, November 2006.
- [95] K. Nakamura, Scott Veirs, Christian P. Sarason, Russell E. McDuff, Fritz Stahr, Dana R. Yoerger, and Albert M. Bradley. Chemical signals in rising buoyant plumes and tidally oscillating plumes at the Main Endeavour vent field, Juan de Fuca Ridge. In *Eos. Trans. AGU*, volume 81, 2000. Fall Meet. Suppl., Abstract OS52I-05.
- [96] Inc. Ocean Marine Industries. Capsum METS methane sensor brochure. 2006, Nov. 12.
- [97] José Ochoa. A practical determination of CTD platinum resistance thermometer response time, and its use to correct salinity bias and spikes. *Deep-Sea Research*, 36(1):139–148, 1989.
- [98] Giuseppe Orilio, Giovanni Ulivi, and Marilena Vendittelli. Fuzzy maps: a new tool for mobile robot perception and planning. *Journal of Robotic Systems*, 14(3):179–197, 1997.

- [99] Giuseppe Orilio, Giovanni Ulivi, and Marilena Vendittelli. Real-time map building and navigation for autonomous robots in unknown environments. *IEEE Transactions on Systems, Man, and Cybernetics–Part B: Cybernetics*, 28(3):316–333, June 1998.
- [100] Daniel Pagac, Eduardo M. Nebot, and Hugh Durrant-Whyte. An evidential approach to probabilistic map-building. In *IEEE International Conference on Robotics and Automation*, pages 745–750, Minneapolis, Minnesota, April 1997.
- [101] Shuo Pang. *Reactive Planning and On-line Mapping for Chemical Plume Tracing*. PhD thesis, Department of Electrical Engineering, University of California, Riverside, 2004.
- [102] Shuo Pang and Jay A. Farrell. Chemical plume source localization. *IEEE Transactions of Systems, Man, and Cybernetics–Part B: Cybernetics*, 36(5):1068–1080, 2006.
- [103] R. K. Pearson. Exploring process data. *Journal of Process Control*, 11:179–194, 2001.
- [104] Ronald K. Pearson. Outliers in process modeling and identification. *IEEE Transactions on Control Systems Technology*, 10(1):55–63, January 2002.
- [105] R. D. Prien, D. P. Connelly, and C. German. In situ chemical analyser for the determination of dissolved fe(ii) and mn(ii). In *ASLO Ocean Sciences Meeting*, Honolulu, HI, 2006. Abstract OS44B-06.
- [106] M. Ribo and A. Pinz. A comparison of three uncertainty calculi for building sonar-based occupancy grids, 2001.
- [107] M. D. Rudnicki and H. Elderfield. Theory applied to the Mid-Atlantic Ridge hydrothermal plumes: the finite-difference approach. *Journal of Volcanology and Geothermal Research*, 50:161–172, 1992.
- [108] M. D. Rudnicki and C. R. German. Temporal variability of the hydrothermal plume above the Kairei vent field, 25° s, Central Indian Ridge. *Geochem. Geophys. Geosys.*, 3(2), 2002. doi:10.1029/2001GC000240.
- [109] Hanumant Singh. *An Entropic Framework for AUV Sensor Modelling*. PhD thesis, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution Joint Program in Oceanography/Applied Ocean Science and Engineering, June 1995.
- [110] K. G. Speer and P. A. Rona. A model of an atlantic and pacific hydrothermal plume. *Journal of Geophysical Research - Oceans*, 94(C5):6213–6220, May 1989.

- [111] Fritz Stahr, Russell McDuff, Dana Yoerger, Albert Bradley, and K. Nakamura. Heat flux measurements at the Main Endeavour vent field, Juan de Fuca Ridge. In *Eos. Trans. AGU*, volume 81, 2000. Fall Meet. Suppl., Abstract OS52I-03.
- [112] Inc. SubChem Systems. Subchempak analyzer. 2006, Nov. 12.
- [113] Richard E. Thomson, Steven F. Milhály, Alexander B. Rabinovich, Russell E. McDuff, Scott R. Veirs, and Frederick R. Stahr. Constrained circulation at Endeavour ridge facilitates colonization by vent larvae. *Nature*, 424(31), 2003.
- [114] W. J. Thomson, R. E. McDuff, F. R. Stahr, D. R. Yoerger, and M. Jakuba. Heat flux from the endeavour segment of the juan de fuca ridge. In *Eos. Trans. AGU*, volume 86, 2005. Fall Meet. Suppl., Abstract T31A-0489.
- [115] S. Thrun. Probabilistic algorithms in robotics. *AI Magazine*, 21(4):93–109, 2000.
- [116] Sebastian Thrun. Learning metric-topological maps for indoor mobile robot navigation. *Artificial Intelligence*, 99(1):21–71, 1998.
- [117] Sebastian Thrun. Learning occupancy grid maps with forward sensor models. *Autonomous Robots*, 15:111–127, 2003.
- [118] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic Robotics*. MIT Press, Cambridge, Massachusetts, 2005.
- [119] J. S. Turner. Turbulent entrainment: the development of the entrainment assumption, and its application to geophysical flows. *Journal of Fluid Mechanics*, 173:431–471, 1986.
- [120] S. Veirs. *Submarine volcanic heat flux and hydrography: observations and models of the Main Endeavour vent field in the Northeast Pacific*. PhD thesis, University of Washington, 2003.
- [121] Scott R. Veirs, Russell E. McDuff, Marvin D. Lilley, and John R. Delaney. Locating hydrothermal vents by detecting buoyant, advected plumes. *Journal of Geophysical Research*, 104(B12):29,239–29,247, December 1999.
- [122] Neil Vickers. Mechanisms of animal navigation in odour plumes. *Biological Bulletin*, 198(2):203–212, 2000.
- [123] Michael R. Wandel, Achim Lilienthal, Tom Duckett, Udo Weimar, and Andreas Zell. Gas distribution in unventilated indoor environments inspected by a mobile robot. In *Proceedings of the IEEE International Conference on Advanced Robotics (ICAR 2003)*, Coimbra, Portugal, 2003.

- [124] M. Weissburg. The fluid dynamical context of chemosensory behaviour. *Biological Bulletin*, 198(2):188–202, 2000.
- [125] Marc J. Weissburg and David B. Dusenbery. Behavioral observations and computer simulations of blue crab movement to a chemical source in a controlled turbulent flow. *Journal of Experimental Biology*, 205:3387–3398, 2002.
- [126] M.J. Weissburg. *Orientation and Communication in Arthropods*, chapter From odor trails to vortex streets: chemo and mechanosensory orientation in turbulent and laminar flows, pages 214–246. Birkhauser, Basel, 1997.
- [127] M. A. Wetzler, J. W. Lavelle, G. A. Cannon, and E. T. Baker. Variability of temperature and currents measured near Pipe Organ hydrothermal vent site. *Marine Geophysical Researches*, 20:505–516, 1998.
- [128] Louis L. Whitcomb, Dana R. Yoerger, and Hanumant Singh. Combined doppler/lbl based navigation of underwater vehicles. In *Proceedings of the 11th International Symposium on Unmanned Untethered Submersible Technology (UUST99)*, Durham, New Hampshire, USA, August 1999.
- [129] J. Scott Willcox, James G. Bellingham, Yanwu Zhang, and Arthur B. Baggeroer. Performance metrics for oceanographic surveys with autonomous underwater vehicles. *IEEE Journal of Oceanic Engineering*, 26(4):711–725, October 2001.
- [130] D. Wolf and S. Sukhatme. Towards mapping dynamic environments, 2003.
- [131] Denis F. Wolf and Gaurav S. Sukhatme. Mobile robot simultaneous localization and mapping in dynamic environments. *Autonomous Robots*, 19(1):53–65, July 2005.
- [132] D. R. Yoerger, A. M. Bradley, F. R. Stahr, and R. E. McDuff. Survey of deep-sea hydrothermal vent plumes with the Autonomous Benthic Explorer (ABE). In *Proceedings of the 12th International Symposium on Unmanned Untethered Submersible Technology (UUST01)*, Durham, New Hampshire, USA, 2001.
- [133] Dana R. Yoerger, Albert M. Bradley, and Barrie B. Walden. The Autonomous Benthic Explorer (ABE): An AUV Optimized for Deep Seafloor Studies. In *Proceedings of the Seventh International Symposium on Unmanned Untethered Submersible Technology (UUST91)*, pages 60–70, Durham, NH, 1991.
- [134] Dana R. Yoerger, Michael Jakuba, Albert M. Bradley, and Brian Bingham. Techniques for deep sea near bottom survey using an autonomous underwater vehicle. *International Journal of Robotics Research*, 2006. To appear.

- [135] Dana R. Yoerger, Pamela G. Murray, and Fritz Stahr. Estimating the vertical velocity of buoyant deep-sea hydrothermal plumes through dynamic analysis of an autonomous vehicle. In *Proceedings of the 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Maui, HI, USA, November 2001.
- [136] Dimitri Zarzhitsky, Diana Spears, David Thayer, and William Spears. Agent-based chemical plume tracing using fluid dynamics. In *Lecture Notes in Computer Science*, volume 3228. Springer-Verlag, 2004.
- [137] Yanwu Zhang, Arthur B. Baggeroer, and James G. Bellingham. Spectral-Feature Classification of Oceanic Processes Using an Autonomous Underwater Vehicle. *IEEE Journal of Oceanic Engineering*, 26(4):726–741, October 2001.
- [138] R. K. Zimmer and C. A. Butman. Chemical signaling processes in the marine environment. *Biological Bulletin*, 198:168–187, 2000.